

Towards a German élite

A new government initiative to promote university development is a welcome attempt to stem the exodus of outstanding researchers from Germany. But it is also necessary to sustain deeper reform of the country's academic system.

Last week an illustrious circle of ten top managers from science and industry assembled in Berlin to meet Chancellor Gerhard Schröder for an 'innovation dinner' in his office. He was hoping for impetus to his goal of enhancing the quality, competitiveness and worldwide recognition of Germany's research and higher-education system.

Whether or not the meeting served this purpose, its widespread publicity demonstrated Schröder's intent to tackle the country's pressing structural problems. That weakness in science and innovation is one of them is no secret but, unlike more obvious difficulties with the social-security systems, science policies seldom make headlines, and debate about reform had been muted.

That was before the phenomenal media coverage of Schröder's announcement, at the Social Democrats' New Year retreat in Weimar, of his intention to make the rusty research system ship-shape, and to create a number of German élite universities. Not surprisingly, the plan received a mixed response — the term 'élite' still stirs resentment in some German intellectuals. It is true that university excellence cannot be created by government decree, but students, researchers and science managers in Germany should be pleased that their needs have become the first major political issue of the year.

But what can Schröder actually do? Establishing a new élite university is the least likely option, and would be misguided — Germany's 100 or so research universities are not as weak as is sometimes alleged. There is no scientific powerhouse equivalent to, say, the Massachusetts Institute of Technology, but quite a few universities — in Berlin, Munich, Heidelberg, Aachen, Tübingen and Göttingen, for example — can readily compete in one field or another.

If Schröder is willing to increase investment in science (and he seems to be serious about it), he should build on existing strengths, rather than try to conjure up a showpiece university out of nothing.

More importantly, he must use his authority in matters of general policy to create a legal and regulatory framework in which research and education can flourish, freed from anachronistic hierarchies and from the bureaucracies of the *Länder* (states).

The main defect of the German university system is that it is top-heavy. Full professors are endowed with veritable kingdoms of power and privilege, whereas most young researchers and PhD students suffer ridiculously subordinate and dependent conditions — the exceptions only prove the rule. Young scientists who struggle through the system, usually on half-time contracts but expected to act as workhorses, risk ending up at 40 knocking on closed doors where, all too often, the academic appointment plans — set up by state officials, not by the universities — make no provision to employ them.

This, rather than the inferiority of German research per se, is the reason for the mass migration across the Atlantic of Germany's brightest minds. Competition in the United States may be even fiercer, but at least it has a culture of science that rewards commitment and merit, and where young investigators are more likely to be treated with due respect.

To be fair, Germany's academic culture is changing. The introduction three years ago of independent junior professorships is one example. Meanwhile, universities in Bavaria have been given some autonomy, and have begun to merge their respective strengths — to the benefit of medical students in Munich, for example. Yet the Humboldt University in Berlin has to cut 80 of 300 faculty positions by 2009, the average student-to-professor ratio is nightmarishly high, the right to select students is an exception rather than the rule, and the introduction of tuition fees is banned by law.

That awareness of these constraints has reached the highest levels of government is welcome. Schröder clearly intends to prioritize the health of academic science. Scientists should seize the moment. ■

Back to the Moon

Despite scientific casualties, sending people to the Moon is politically the best way forward for science in the long term.

President Bush's plan to steer NASA towards the Moon and Mars is, for now, more talk than action. The lunar landings, and most of the spending, will come long after he leaves office. Let's accept that his motives go beyond election-year grandstanding, however. Did he make the right decision?

Yes, with reservations. Having decided correctly to ground the shuttle and scale back investment on the space station, Bush chose — probably with an eye on US public opinion — to keep sending astronauts into space. The alternative was to cut NASA back to being a research-only agency. Scientists were fooling themselves if they thought that the money saved from the space shuttle would go on space telescopes. It would simply have disappeared.

The Moon is a useful, nearby place to find out whether human spaceflight still has a viable future. Arguments based on destiny or the 'human spirit' only go so far. In the end, astronauts need something to do out there. A small, 'tended' (rather than permanently occupied)

lunar base can help settle the old, but never fully resolved, debate over sending humans or robots into space. The two will work together at first, just as in a modern factory on Earth. We can hope that NASA's twenty-first-century lunar exploration will advance the science of telerobotics in a meaningful way. If astronauts and machines are assigned work based solely on their merits and cost-effectiveness, we may see human visits to the base become less frequent as engineers learn to achieve the same tasks remotely. This evolution could be one of the new Moon programme's most valuable products, and could change the way we approach Mars exploration. Maybe when the time comes to send humans to Mars, we'll find that we don't need to.

But to reach that point, NASA should not prejudice the new Moon programme to favour astronauts over their robotic assistants. This will require clarity of purpose and honest decisions about engineering and funding priorities. Bush's plan calls on NASA to reinvent itself and break some ingrained bad habits. A tall order, but a good one. ■





California reeling
University institutes
dogged by shortfalls
in cash
p274



Close call
Study results alter
pecking order for
chicken welfare
p275



On the brink
French researchers
threaten resignation
over science funds
p276



Pet project
US authorities
prepare to clamp
down on animal trade
p277

NASA's drive to revisit the Moon leaves no scope for Hubble

Tony Reichhardt, Washington

President George Bush's call last week to return US astronauts to the Moon by 2020 has won him plaudits from the space community. But on 16 January his plan claimed its first high-profile victim when NASA administrator Sean O'Keefe announced that the Hubble Space Telescope will be retired about three years early.

Bush's proposal, which calls for a manned lunar landing sometime between 2015 and 2020, requires the most sweeping overhaul in NASA's history. Researchers contacted by *Nature* say that with about \$12 billion earmarked for the project over the next five years, the space agency should have enough money at least to get started on it.

But of this money, Bush's plan includes only \$1 billion in additional funding, so the space shuttle will be phased out by 2010. Together with other savings, including drastic cuts in research on the still unfinished International Space Station, the agency should save an additional \$11 billion over the next five years to apply to the Moon programme.

Hubble will be an indirect victim of this cost-cutting. NASA decided in 1997 to extend the telescope's life from 2005 to 2010, and astronomers had been lobbying for another extension (see *Nature* 424, 603; 2003). But O'Keefe told a glum gathering of Hubble researchers and engineers at the Goddard Space Flight Center in Greenbelt, Maryland, that the shuttle will no longer be allowed to visit the telescope, in effect cancelling a 2006 visit by astronauts to add new instruments to the telescope. Observations will probably end sometime after 2007, as its parts begin to fail. The details of Hubble's demise have not been finalized, but a robotic craft is likely to bring it back into Earth's atmosphere, where it would burn up.

O'Keefe said that his decision was based on several factors, including the availability of ground-based telescopes for optical astronomy. But money figured prominently in his thinking. To satisfy recommendations made by the board that investigated the loss of the space shuttle Columbia last February, NASA will have to develop the capability to



The Hubble Space Telescope will meet a fiery end when it hits Earth's atmosphere sometime after 2007.

inspect shuttles in orbit. That job is relatively easy to do when the shuttle is near the space station, but is more expensive near Hubble, which orbits farther from Earth. Scrapping the 2006 servicing trip to Hubble will avoid these costs, but will also shorten the telescope's lifetime.

Astronomers who use Hubble are not the only group to lose out under Bush's plan. The space station will be complete by 2010, and NASA will fund research there until 2017. But work in areas not related to human spaceflight, such as materials science, is expected to be greatly reduced or scrapped altogether.

James Pawelczyk, a physiologist at Pennsylvania State University who served on a 2002 panel to prioritize station science, says that the 2017 date still leaves "a fair amount of time to do research", even though key pieces of laboratory equipment will not arrive until 2008 at the earliest. And it is not yet clear, he adds, whether experiments in areas such as the effects of gravity on plants and animals will be considered relevant to the new mandate on human exploration.

Other criticisms have come from sup-

porters of space exploration, such as Senator Bill Nelson (Democrat, Florida), a one-time shuttle passenger, who questioned whether funding for the programme will be adequate. But experts contacted by *Nature* were saying that the numbers seem to add up. Michael Griffin, a former head of NASA's exploration office, says that with NASA management reforms and clever engineering, the \$12 billion it will have for the first five years "is probably okay".

NASA gave no estimate of the final cost to land people on the Moon, although it would certainly be tens of billions of dollars. Griffin, who now directs In-Q-Tel, a technology firm based in Arlington, Virginia, estimated in congressional testimony last October that NASA would need an additional \$5 billion beyond its annual \$15-billion budget to establish a lunar base within a decade.

If nothing else, Pawelczyk says that the emphasis on human exploration has given the agency a much-needed new focus. "The entire space community has been paralysed" in debates over where to go next, he points out. Bush's decision fixes that in one stroke, he says. "Now it's up to NASA." ■

Bird flu sparks worldwide bid to prevent human pandemic

Alison Abbott and David Cyranoski

The outbreak of highly contagious bird flu that is sweeping Asia, and which has killed at least five people, has put the World Health Organization (WHO) on full alert.

WHO scientists are now pulling out all the stops to trace the origin of the virus, known as H5N1, and to develop vaccine candidates. The spectre of a new human influenza pandemic, similar to that which killed millions in 1918, looms over their effort. Pandemics arise when bird or pig flu viruses jump the species barrier to humans and mutate into highly contagious forms.

The current bird flu outbreak, which has wreaked havoc among poultry in Vietnam, Japan and South Korea, is exceptionally worrying, says Albert Osterhaus, an influenza expert at Erasmus University in Rotterdam, the Netherlands. Bird flu outbreaks in the Netherlands and Hong Kong last winter were rapidly contained and led to the deaths of only three people (see *Nature* 422, 247; 2003). But Osterhaus says that management of the current outbreak has been loose and surveillance inadequate. There is no sign that the outbreak is abating, and it may already have

spread to birds in Cambodia or Laos, where surveillance is also lax.

Viruses that jump the species barrier tend to be highly pathogenic, as the new host's immune system has had no time to build up defences. Although dangerous, the viruses are not necessarily contagious because they are adapted to survive in their original host. But the greater the number of people who catch the flu, the greater the chance that the virus will mutate into a form that is easily transmitted between humans. As yet, human-to-human transmission does not seem to have occurred.

Scientists in Japan, Hong Kong and the United States have sequenced the genomes of viruses found in birds in Vietnam, Japan and Korea, as well as those isolated from Vietnamese patients who have died.

The genomic data could reveal where the virus first emerged, its path of transmission and mutation rate, and how to control the spread, says Masato Tashiro, a virologist at the National Institute of Infectious Diseases in Tokyo. "The strains are similar to that isolated from a 2003 Hong Kong victim," he adds.

The data are being pooled in the Influenza Sequence Database based in Los Alamos, Cali-



No escape: chickens are bagged up prior to burial in an effort to halt the spread of bird flu.

fornia, to aid development of human vaccine candidates. One candidate, made for the 2003 Hong Kong outbreak, is already being tested and if shown to be effective against the current strain will offer the effort a head start. "Should the few cases we are currently seeing in Vietnam turn out to signal a pandemic, we shall be ready for it," says Maria Cheng, a WHO spokeswoman based in Geneva. ■

Director quits as cash cutbacks hit Californian labs

Rex Dalton, San Diego

The California Institutes for Science and Innovation — four high-profile interdisciplinary facilities built at a cost of \$400 million — are going through a turbulent time just months before researchers move into the new buildings.

A bleak financial outlook, caused by fiscal difficulties at state level, is widely believed to have contributed to the departure of Marvin Cassman, head of one of the new facilities — the Institute for Bioengineering, Biotechnology and Quantitative Biomedical Research (QB3). Cassman, a former director of the US National Institute of General Medical Sciences who was appointed with much fanfare in 2002, quietly quit late last month and did not respond to requests for an interview.

The institutes are designed to channel expertise from across the University of California (UC) campuses into scientific discoveries that will spur statewide economic development. They were to have had an annual budget of at least \$10 million



Shaky start: the San Francisco labs of the University of California's QB3 Institute, which has lost its inaugural director.

each for operating costs. But California is going through a fiscal crisis. The proposed budget for the next financial year, which will be finalized by July, cuts \$370 million from the total UC budget and hands the centres just \$1 million each.

Because scientists at the institutes will fund their research through external grants,

the researchers say they are confident that good science will still get done at the new facilities. But many fear that their efforts could be held back by problems with operating costs. Keith Yamamoto, a biochemist at UC San Francisco and executive vice-dean of the university's medical school, says the issue is of "grave concern".

At the California Institute for Telecommunications and Information Technology, based at the UC San Diego and Irvine campuses, director Larry Smarr says that they are "slowing down the growth" of programmes and may not initially occupy the entire new San Diego building. Administrators at the campuses that host the other two centres — the California NanoSystems Institute at Los Angeles and Santa

Barbara and the Center for Information Technology Research in the Interest of Society at four campuses — say that they will use university funds to pay the operating costs.

QB3 will be run until 30 June by Graham Fleming, director of physical biosciences at UC Berkeley. Officials are exploring options for QB3's directorship after then. ■

Feathers fly over welfare of hemmed-in hens

Laura Nelson, London

Cramming chickens together in a confined living space isn't as bad for them as one might think, according to a study published on page 342 of this issue.

The results, which reveal that temperature and air quality have a greater impact on chicken mortality than cramped conditions, are expected to be used by the poultry industry to attack proposed regulations by the European Union (EU) on chicken welfare. But defenders of the regulations maintain that controlling stocking density is the most realistic way to improve the welfare of the chickens.

Worldwide, more than 20 billion chickens are killed for human consumption every year, and the industry is facing growing criti-

cism over the conditions in which most of the birds are kept.

The researchers, led by Marian Dawkins, an animal behavioural scientist at the University of Oxford, say that their study is the largest attempt so far to shed some scientific light on this emotionally charged debate. "Nobody had looked at this before," says Dawkins.

The three-year study, which was funded by the UK government, monitored 2.7 million birds reared by ten broiler-producing companies. The birds were kept at densities of between 30 kg per square metre — the maximum proposed in a 2000 report that is expected to form the basis of the new EU regulations — and 46 kg per square metre. Welfare was assessed by measuring mortality, levels of the stress hormone corticosterone

in the faeces, ease of walking and the presence of skin lesions on the birds' legs and feet.

Although chickens reared in the more crowded conditions grew more slowly, the number dying, being culled as unfit, or showing leg injuries did not correlate directly with stocking density, the authors found. But mortality was directly related to humidity and temperature. Disparities between the birds held by different producers had far more effect on mortality than did stocking density.

"The results showed that although very high stocking densities do affect chicken welfare, stocking density per se is, within limits, less important than other factors in the birds' environment," says Dawkins.

But Britain's Royal Society for the Prevention of Cruelty to Animals (RSPCA) says that the study doesn't sufficiently emphasize adverse impacts other than mortality. Caroline Le Sueur, an RSPCA spokeswoman, points out that the results show that birds grown in crowded conditions are more likely to jostle one another and to walk irregularly.

The study should have dwelt more on the "quality of life" of closely confined animals, Le Sueur suggests. "We would be very concerned if industry thought it could get away with high stocking densities."

Defenders of the EU regulations, which are expected to be published later this year, point out that stocking density is easier for inspectors to monitor than factors that directly correlate with chicken mortality, such as the levels of ammonia in the air the birds breathe.

But European poultry producers say the study shows that stocking density is the wrong focus for the regulations. A spokesman for the British Poultry Council says the industry recognizes that good chicken welfare is good for business, but he adds that US producers don't have to deal with welfare rules for chickens. ■



Close encounters: temperature and air quality may be the main influences on the health of chickens.

Ministers prepare to back neuroscience network

Declan Butler, Paris

Brain scientists around the world wrestling with the thorny problem of how to share their data should soon find their lives getting a little easier.

Plans for an international facility to integrate such information are expected to be endorsed next week at a meeting in Paris of 30 science ministers from countries in the Organisation for Economic Co-operation and Development (OECD).

The planned International Neuroinformatics Coordinating Facility would bring together disparate neuroscience databases to allow researchers to share and compare their results more readily

(see *Nature* 406, 822–825; 2000).

The facility would build its own databases, nurture agreements between nations on data sharing, and even fund some research on neuroinformatics, says Stefan Michalowski, executive secretary of the OECD's Global Science Forum. The annual cost of the administrative body that would run the facility would be about US\$1.2 million.

The facility would be modelled on the Copenhagen-based Global Biodiversity Information Facility, which performs a similar function with biodiversity data and was endorsed at the last OECD ministerial science meeting in Paris in 1999.

The ministers meeting in Paris are also

expected to declare their support for a next-generation electron–positron linear collider to succeed the Large Hadron Collider, which is currently being assembled at CERN, the European particle-physics lab near Geneva. They are also likely to call on governments and research agencies to have the new collider built as a genuinely global project.

In addition, the ministers are set to draft a declaration supporting open access to data generated by publicly funded research. The resolution would ask the OECD to develop international principles and guidelines to help maximize open access, for example by making data sharing a requirement in grant proposals. ■

Resignation threats add steel to French revolt

Declan Butler, Paris

Axel Kahn's is one of the most well-heeled genetics labs in France. As a pillar of the research establishment, he is not given to tempestuous announcements. But earlier this month, he threatened to resign as director of the Cochin Institute in Paris unless the government takes immediate and significant steps to boost French research.

Kahn, together with many of the country's top scientists, is a prominent member of a researchers' revolt. The protesters, known as the Save Research movement, made their demands on 7 January in an online petition. They want the government to pay research agencies the millions of euros it owes them, to boost recruitment of young scientists, and to launch a national debate on science. To give teeth to their demands, signatories pledge to resign all administrative duties, an act that could bring the country's research system to its knees.

The seeds of revolt were sown last year, when the national research agencies bore the brunt of a 1.3% cut in the country's €8.8-billion (US\$10-billion) research budget. At Kahn's laboratory, for example, public funding fell 10%, and staff numbers dropped by



Axel Kahn, director of the Cochin Institute, says he will resign as part of a protest by French scientists.

15 to around 150. "And we are one of the top labs in France," he says. "Just imagine the situation elsewhere."

Faced with crushing public deficits, the finance ministry has also withheld millions from the agencies, bringing them close to bankruptcy. The CNRS, France's national research agency, was until last week owed €172 million. The agency was only able to avoid going broke after the government freed up €103 million, promising to repay the rest in 2005.

But it is the job situation that sparked the protests. At INSERM, the national biomedical research agency, staff numbers are being cut and only 30 new positions are available this year, compared with 90 in 2003. New three- to five-year contracts, paying €1,900 a month, are also being shunned by young scientists, as it is almost impossible to rent an apartment or get a loan on a short-term contract.

Kahn says that anger among PhD students infected more senior staff, leading them to establish the petition, which has attracted more than 17,000 signatures. As a well-known public figure, Kahn feels obliged to go one step further than other signatories and resign his research position. He seems to be deadly serious about carrying this through: "I can't allow myself to make threats that are not credible."

Last weekend, science minister Claudie Haigneré received Alain Trautmann, co-head of the biology department at the Cochin Institute and spokesman for the Save Research movement. Haigneré described the meeting as "constructive" and promised an inter-ministerial meeting to discuss the demands. ■

♦ <http://recherche-en-danger.apinc.org>

Europe's researchers up in arms over clinical-trial rules

Alison Abbott

Some 2,000 clinical researchers have signed an Internet petition calling for changes to proposed new European Union (EU) rules on patient trials. They say that, as the revised rules stand, they would place immense bureaucratic burdens on their backs.

Under the new rules, researchers who conduct clinical trials would have to fill out mountains of forms, follow and report the well-being of patients to a central database, and accept full liability for trials whose responsibility was previously shared among different authorities.

The researchers admit that they have woken up late to the problems posed by the 2001 EU Directive on Clinical Trials, which is due to be implemented through national laws by this May. But they are hopeful that its impact will be curtailed by a more detailed, follow-up directive now being drafted in Brussels, and through the discretion of national governments in passing the laws.

The directive was originally drawn up in the Enterprise Directorate of the European Commission to harmonize the regulation of large clinical trials conducted by the drug industry. But as it developed, its remit was extended to cover all patient trials, including

smaller ones conducted by academics involving drugs already on the market.

"All along, there has been too little consultation with those affected by the rules," says Brian Moulton, a pharmacologist at the Dublin-based Irish Clinical Oncology Research Group, which helped to organize the petition. "But we hope that the extraordinary level of support for our petition will influence policy-makers."

The directive requires clinical researchers to conform with two detailed sets of EU guidelines, called Good Manufacturing Practice and Good Clinical Practice, and to follow new procedures designed to control the large-scale clinical trials undertaken by drug companies. Scientists hope to relax these obligations by influencing the contents of a subsidiary set of rules on clinical research being drawn up by the commission.

A spokeswoman for the commission says that it is now sensitive to academics' concerns, which have been supported by research commissioner Philippe Busquin. Commission officials are in "internal debates about the extent to which academic research should be tied to the main directive," she says.

Critics say that, unless the guidelines are worded appropriately, trials that the drug

industry doesn't want to run could be endangered because academics could no longer afford to do them. Such trials include efficacy comparisons of drugs that are already in widespread use, and tests of procedures using surgery or medical devices.

Peter Boyle, director of the International Agency for Research on Cancer in Lyon, France, cites a recent 20-year cancer study as a potential victim of the rules. The Milan-based European Institute of Oncology's comparison of radical mastectomy against radical lumpectomy for breast cancer, he points out, "showed no difference in outcome between the procedures, so it had a huge impact for millions of women — but it's not likely that it could be done under the new rules because the administration costs would be too high."

But the problem doesn't just relate to cost. Jane Armitage, a clinician at the University of Oxford's Clinical Trial Service Unit, UK, says that under the new rules, many trials that academics do now could not take place, because only drugs explicitly labelled for use in a trial could be used. "Many trials are based on doctors recommending patients to take — or avoid — aspirin, or other drugs that they do not provide themselves," she says. ■

♦ www.saveeuropeanresearch.org

Health concerns prompt US review of exotic-pet trade

Erika Check, Washington

For years, biologists have railed against the US\$6-billion international wildlife trade, claiming that it has driven a long list of species to the brink of extinction. Now the US government seems to be paying attention to such protests — but not because of any newfound concern for animal welfare.

Instead, US officials are worried about diseases such as severe acute respiratory syndrome (SARS) and monkeypox, which caused epidemics last year when they jumped from animals to people. Conservation groups have been arguing that the animal trade is to blame for such epidemics — and they seem to be gaining the ear of the government.

The latest call came on 14 January, when the New York-based group Wildlife Trust convened biologists and ecologists from ten countries in Bangalore, India. The scientists issued a declaration calling for “an urgent reassessment of current wildlife trade practices”.

Last summer, a coalition of US epidemiologists and veterinarians said that the federal government should expand an order that restricted the importation of rats and prairie dogs into the United States. The animals were responsible for an outbreak of monkeypox that infected 40 people in the midwest last summer.

“To protect public health, this order should be permanently sustained and expanded to restrict the movement of exotic wildlife with potential adverse impact on public health,” the Council of State and Territorial Epidemiologists and the National Association of State Public Health Veterinarians wrote in a letter to government agencies including the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia.

Now there are signs that such arguments are hitting home. On 13 January, the US health department declared the country off-limits to civets in a precaution against an outbreak of SARS, the flu-like illness first reported in China last February. Nobody has yet proved that the animals were the source of the SARS epidemic, but scientists have found that civets on sale in Chinese markets harbour viruses that are very similar to those that cause SARS. US health officials say that since the



Caught out: this US family found themselves in quarantine last summer after one of their pet prairie dogs gave them monkeypox.

civet market is not large in the United States — only 50 are imported every year — this limited evidence convinced them to enact a ban. And they say that they are looking for a more widespread approach to the problem of animal-borne, or zoonotic, diseases.

“There’s been a lot of increased discussion from federal agencies and independent groups to see what can be done to reduce the risk of importing zoonotic disease into this country,” says Paul Arguin, a specialist in animal-borne diseases at the CDC. Arguin says that the CDC has been talking with the public-health veterinarians’ group on ways to address this threat. “It’s fair to say that the CDC is in support of increased regulation of the exotic-pet trade,” he says.

Mary Pearl, president of the Wildlife Trust, says that such regulation is overdue. For a start, she says, governments should use better quarantine and surveillance methods for animals coming into the country, and could give more funding to enforcement agencies that are supposed to police the illegal wildlife trade.

“They’ve been rapidly closing the barn door after the horse has bolted,” Pearl says. “It would make more sense to have a rational set of policies that requires adequate surveillance of the wildlife trade, because there simply are not enough safeguards to protect public health.”

Telepathic charm seduces audience at paranormal debate

John Whitfield, London

Scientists tend to steer clear of public debates with advocates of the paranormal. And judging from the response of a London audience to a rare example of such a head-to-head conflict last week, they are wise to do so.

Lewis Wolpert, a developmental biologist at University College London, made the case against the existence of telepathy at a debate at the Royal Society of Arts (RSA) in London on 15 January. Rupert Sheldrake, a former biochemist and plant physiologist at the University of Cambridge who has taken up parapsychology, argued in its favour. And most of the 200-strong audience seemed to agree with him.

Wolpert is one of Britain’s best-known public spokesmen for science. But few members of the audience seemed to be swayed by his arguments.

Sheldrake, who moved beyond the scientific pale in the early 1980s by claiming that ideas and forms can spread by a mysterious force he called morphic resonance, kicked off the debate.

He presented the results of tests of extrasensory perception, together with his own research on whether people know who is going to phone or e-mail them, on whether dogs know when their owners are coming home, and on the allegedly telepathic bond between a New York woman and her parrot. “Billions of perfectly rational people believe that they have had these experiences,” he said.

Wolpert countered that telepathy was “pathological science”, based on tiny, unrepeatable effects backed up by fantastic theories and an ad hoc response to criticism. “The blunt fact is that there’s no persuasive evidence for it,” he said. “An open mind is a very bad thing — everything falls out.”

For Ann Blaber, who works in children’s music and was undecided on the subject, Sheldrake was the more convincing. “You can’t just dismiss all the evidence for telepathy out of hand,” she said. Her view was reflected by many in the audience, who variously accused Wolpert of “not knowing the evidence” and being “unscientific”.

In staging the debate, the RSA joins a growing list of London organizations taking a novel approach to science communication (see *Nature* 426, 6; 2003). “We want to provide a platform for controversial subjects,” says Liz Winder, head of lectures at the RSA.

China's human-cloning policy fudges law on cross-species fusions

Tokyo China has outlawed human reproductive cloning but has not legislated for or against controversial pioneering work that fuses human cells with eggs of other species.

Guidelines issued by the Chinese Ministry of Science and Technology and the Ministry of Health ban the cross-species fusion of eggs and sperm. But they make no mention of techniques used by Huizhen Sheng of Shanghai Second Medical University to reprogramme human cells by fusing them with rabbit eggs emptied of their genetic material (see *Nature* **424**, 711; 2003).

The legislation, announced last week, permits the derivation of embryonic-stem-cell lines from surplus embryos left over from *in vitro* fertilization and embryos created by transferring adult cells into empty eggs — a process known as therapeutic cloning or somatic-cell nuclear transplantation. But such embryos cannot be grown beyond 14 days.

Scientific advisers to the Chinese government say that the guidelines are based on the relatively liberal UK legislation.

The announcement coincides with the public but unsubstantiated claims in London by human-cloning advocate Panos Zavos that he has implanted a cloned human embryo into a woman.

NIH defends grants to study sexual behaviour

Washington US National Institutes of Health (NIH) chief Elias Zerhouni has spoken out in defence of almost 200 grants that were criticized by a conservative lobbying group. The grants, which deal with issues such as homosexuality, drug use and sexual behaviour, were put on a 'hit list' by the Traditional Values Coalition. The list was presented to a Congressional committee that oversees the NIH's activities.

Last week, Zerhouni told an NIH advisory committee that his agency had reviewed all the grants on the list, and found compelling reasons to fund all of the studies. He added that he would spell out to Congress the scientific rationale for the studies.

Scientists were pleased that Zerhouni had come to their defence. "It's very easy when individual studies are taken out of context to question their importance or value," says infectious-disease specialist Daniel Kuritzkes, vice-chair of the board of directors of the HIV Medicine Association.

Japan digs deep to find Expo's mammoth star

Tokyo A frozen woolly mammoth is expected to form the centrepiece of an exhibition at the Expo 2005, to be held in Aichi, Japan. This would be the first time ever that a frozen adult mammoth will have been displayed.

The head of a mammoth has been found in the republic of Sakha, northern Russia, and is being stored in liquid nitrogen. The rest of the mammoth is thought to be intact, and the Expo's organizers and the Sakha government has agreed for it to be excavated. A refrigerated research lab would



Shinji Fukukawa heads the committee that aims to bring a frozen mammoth to Expo 2005.

TORU YAMANAKA/AFP/GETTY IMAGES

then be set up at the exhibition to study samples of the mammoth's body parts.

"We want to look at the evolution of the function of various organs," says Naoki Suzuki, a medical-imaging specialist and palaeontologist at the Jikei University School of Medicine in Tokyo, who will supervise the scientific study. Analysis of the mammoth should provide clues as to how it ate and moved. Suzuki also hopes to find out why the mammoth became extinct.

German ruling threatens to uproot transgenic trials

Munich The future of field trials of genetically modified (GM) crops in Germany is in jeopardy, following a ruling that hands approvals of proposed trials to an agency opposed to GM technology.

The 320 field trials so far approved in Germany, for crops such as potatoes and soya, have been in the hands of the environment ministry and the Robert Koch Institute, which belongs to the health ministry. But at the end of January the environment ministry will lose its jurisdiction to the Federal Agency for Nature Conservation, which is hostile to GM agriculture.

Scientists say they are unhappy with the change. "It is an ideological move that is not supportive to basic research," says Gerhard

Swamp census looks in mud for tiger feet

New Delhi Conservationists in India and Bangladesh are set to carry out a large-scale census of Royal Bengal tigers in the Sundarbans, the world's largest mangrove forest.

The number of tigers in the Sundarbans has fallen in recent years as a result of hunting and damage to their habitat by logging. But there has been no accurate census of their population in the entire region, which straddles the border of the two countries. The new count relies on taking casts of tiger paw prints left in the mangrove mud at low tide (right).

Funding for the project, which will also look into the breeding habits of the animals, will come from the United Nations Development Programme.



Wenzel, who works on GM plant breeding at the Technical University of Munich.

DNA-theft bill stores up trouble for research

London Britain is threatening to ban studies of DNA from stored human tissue samples, if a bill debated last week in parliament is passed. The bill sets out principles of consent regarding use of human tissues and makes it an offence to use tissues without prior permission from the donor.

The Department of Health says the bill is

intended to prevent theft of DNA, such as for bogus insurance claims. But critics argue that the confusing wording bans any use of DNA without consent. "The solution is to reword the clause to allow research and prohibit theft," says Richard Sullivan, head of clinical programmes at Cancer Research UK, who worries about the effect on cancer studies.

The bill is an attempt to restore public confidence after scandals at Liverpool's Alder Hey Hospital, from which numerous tissues were taken from dead patients without their relatives' consent (see *Nature* **409**, 655; 2001).

Star quality

Japan's Nobel laureates have a celebrity status that outstrips anything seen by their contemporaries in North America or Europe. How have they balanced the desire to be a positive influence with the need to retain some privacy? David Cyranoski finds out.

Some three decades after sharing the Nobel Prize in Physics for his work on tunnelling by electrons in semiconductors, Leona Esaki is still a hot property. Caught after giving a talk in Tokyo in 2000, he politely declined *Nature's* request for an interview. "Maybe in a couple of months," he apologized. "I'm just too busy." A gruelling schedule of lectures and social engagements, combined with his role as president of the Shibaura Institute of Technology, left him with precious little spare time. "It's the duty of a Nobel laureate in Japan," Esaki explained.

At least today, there are a few others to share that duty. Between 2000 and 2002, four Japanese scientists received Nobel prizes — just one less than the country's researchers gained over the previous century. Remarkably, this followed hard on the heels of an official government target to win 30 Nobel prizes over the next 50 years (see *Nature* **413**, 560–564; 2001).

It seems unlikely that there's a direct connection, but the wave of recognition from Stockholm has resulted in a frenzy of media interest. The four new laureates are household names, and the Japanese public can't seem to learn enough about their personal lives — their hobbies, dietary preferences and even family life have become fodder for magazine articles and television programmes.

The tide of enthusiasm has been a mixed blessing, the scientists say. The Nobel awards gave a much-needed boost to a nation that was engaged in a bout of soul-searching in the midst of an extended economic recession. More specifically, scientists believe that the publicity has helped to protect research budgets from cuts, and has reignited interest among students who had seemed increas-



Eye of the storm: Japan's four recent Nobel laureates (clockwise from above) Hideki Shirakawa, Ryoji Noyori, Koichi Tanaka and Masatoshi Koshiba have become the centre of media attention since their awards.

ingly averse to taking courses in the sciences. "Overall, it has been very positive," says Ryoji Noyori of Nagoya University, who shared the 2001 chemistry prize for developing a method to preferentially produce molecules whose structures have either left- or right-handed orientations. But when researchers start to be treated as entertainers, he warns, there is a price to pay. "That kind of treatment may spoil the image of science."

Feeding frenzy

The four recent Nobellists were ill-prepared for a whirlwind of media interest that often crossed the line into pure entertainment. One New Year's TV special in 2003 featured Masatoshi Koshiba of the University of Tokyo — who shared the 2002 physics prize for confirming the existence of elusive subatomic particles called neutrinos — and his wife. The show was billed as tackling Koshiba's "hardships, his surprising favourite dishes, and what he's like around the house". Such excursions into Koshiba's private life had by then become commonplace. "There were so many programmes that I was on, I don't remember which one you are talking about," he says.

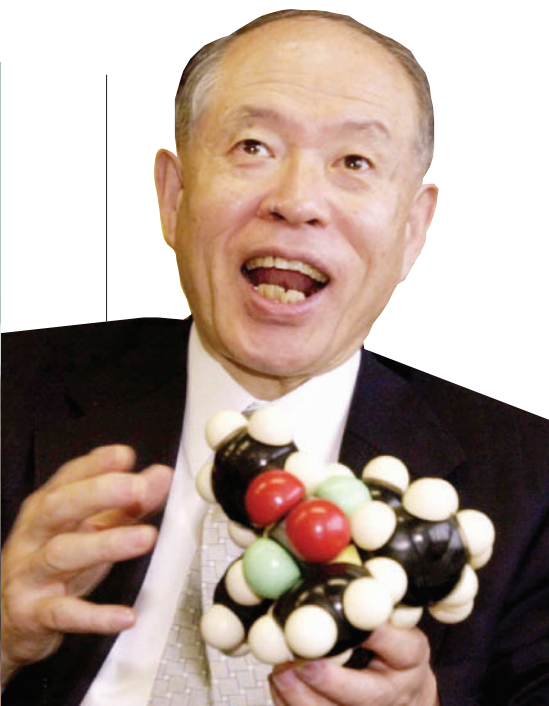
Koichi Tanaka of the Shimadzu Corporation in Kyoto, who shared the 2002 chemistry prize for developing a mass-spectrometry method to study large biological molecules, had much the same experience. "They wanted to know everything about how I lived and what I normally ate," he says. "Nobel winners



are celebrated like sports stars, like they are a completely different type of person."

Even within this small band of celebrities, Tanaka became a particular favourite of TV producers and magazine editors. A previously unknown industrial scientist, he was presented as a model of endeavour and modesty — a down-to-earth star in a world of flashy celebrities. In December 2002, for example, a popular weekly magazine latched on to the rumour that Tanaka met his wife through the *omiai* system — a traditional arrangement in which a mediator brings couples together. The article documented the regrets of women who had passed up the 'Tanaka-type' male on the *omiai* scene. If such men's charms were appreciated, it suggested, more successful marriages would result.

The cult of celebrity that surrounds Nobel laureates in Japan isn't restricted to



the country's own. John Walker, director of the UK Medical Research Council's Dunn Human Nutrition Unit in Cambridge, who shared the 1997 chemistry Nobel for demonstrating how cells make the energy-storing molecule adenosine triphosphate, was surprised to find that his lectures in Japan were televised. This was in stark contrast to the attention given by the British media to the nation's Nobel laureates. "None of us have been accorded celebrity status," Walker observes. "People did comment on the acclaim given to the Spice Girls compared with that given to me."

But sometimes, what Japan's Nobel prizewinners really, really want is the anonymity that Walker still enjoys. Noyori, for instance, became embarrassed by media references to his taste for wine — sometimes linked to his understanding of delicate

chemical processes. "I like to have wine sometimes," he says. "But I'm no expert." Tanaka also came to despair of the way in which he was portrayed. "I was treated like a TV personality," he says. "But they get paid to put their private lives on display. That's not what I am. I started to wonder: 'Why do I have to get my picture taken?'"

Commercial interests also jumped on the Nobel bandwagon. After mentioning in interviews that he liked *konbu* — a type of edible seaweed — and certain beverages, packages containing the products turned up at Tanaka's door. "If it was from a company, I couldn't write back to thank them," he says. "It would get used as an endorsement for their product."

Enough is enough

Tanaka eventually started refusing any bookings that didn't specifically have to do with building support for science and technology. And Noyori was so distressed by one TV variety show's treatment of Tanaka that he called the station up to complain: "I wanted them to respect his effort and his achievement."

Despite these unsettling experiences, the four laureates are pleased with the support that they've been able to attract for science. Rather than sating a population that had been starved of Nobels, the success made Japan hungry for more. Suddenly, everyone was talking about how to make Japanese science even stronger.

One effect was that three of the prize-winning researchers were themselves catapulted into positions of power. Hideki Shirakawa was getting ready for a quiet retirement from the University of Tsukuba in summer 2000 until he shared the chemistry Nobel for his part in discovering that plastics can conduct electricity. When *Nature* visited him in May 2001, he had an office in the middle of Japan's central government buildings in Tokyo, where he had monthly meetings with the prime minister as a member of the country's highest science policy-making body. "I didn't expect to be here," he admitted.

In October 2003, Noyori became the president of the Institute of Physical and Chemical Research, or RIKEN, which administers several of Japan's biggest national research labs. Earlier in the year, Tanaka assumed the directorship of a laboratory created by his company, and was given a first-year budget of ¥200 million (US\$1.9 million). Tanaka is still a bit shy about its name: the Koichi Tanaka Mass Spectrometry Research Laboratory. "I drop the first part when I refer to it," he says.

From these positions of influence, and with unprecedented access to the media, the laureates promoted the need to give more

support and independence to younger scientists — long recognized by leading Japanese researchers as the main impediment to innovative work in the country's labs. Noyori, who was a regular on policy-making committees even before his award, has no doubt that the interest generated by the Nobels helped science budgets to gain small increases even as overall government spending decreased.

Industrial researchers were particularly delighted by Tanaka's award. Generously funded during the economic boom years, the country's corporate labs have more recently fallen into a spiral of lay-offs and shrinking budgets — with little recognition that many industrial scientists remain capable of genuine innovation. "What surprised me the most is that people didn't think that creative scientists existed in industry," says Tanaka. "It was like the announcement of a new species. I wanted people to understand the value of hard work, of the people behind the scenes."

That message seems to have got across. "It's been quite an inspiration, especially for the younger researchers," says Jun'ichi Sone, general manager of NEC's Fundamental Research Laboratories in Tsukuba, north of Tokyo. He hopes that the award will "shine a light on some of the other great industrial research in Japan".

Perhaps the most satisfying aspect of the publicity generated

by the Nobel successes has been the resurgence of interest in studying science among young Japanese. Koshiba was so committed to the goal of educating the next generation of prizewinning scientists that he put his entire bundle of Nobel prize money — ¥40 million — into a foundation geared to develop science teachers and teaching materials. Hamamatsu Photonics, the company that made the photomultiplier tubes that detected the elusive neutrinos, added ¥60 million, and further donations have come in. "There is wide support from all over Japan," Koshiba says.

More importantly, a general change in attitude may be under way. For years, surveys have revealed declining interest among young Japanese in studying science and mathematics. It's too early to say whether the country's outbreak of Nobel fever can reverse that trend, but one informal poll has provided an encouraging sign. Each year since 1989, the Dai-ichi Mutual Life Insurance Company has asked a sample of Japanese primary schoolchildren what they want to do when they grow up. In the latest poll, a career in academic research was the top choice among boys for the first time — beating even the dream of becoming a star of baseball or soccer. ■

David Cyranoski is *Nature's* Asian-Pacific correspondent.



Gut reaction

Consumers are stocking up on live yoghurts and fermented drinks that claim to improve health. But is there any science behind the marketing of these 'probiotic' products? Alison Abbott investigates.

Glenn Gibson's wife prefers him to tell dinner-party guests that he works as a painter and decorator. That's understandable, because if he talks about his real job as a researcher of gut bacteria at the University of Reading, UK, the conversation all too easily turns to the source of his research material — human excrement.

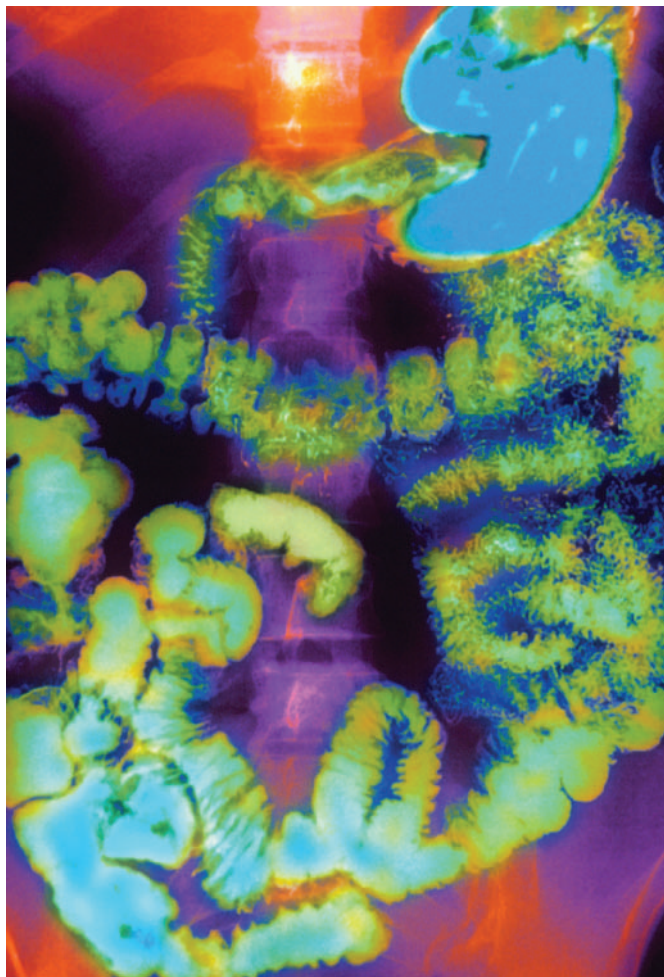
For better or worse, faeces provide the best window into the microbial life of the human gut, a subject that is attracting more funding now than ever before. Partly in reaction to commercial claims that the bacteria in some yoghurts and other 'probiotic' products can boost health, the European Union (EU) has invested more than €15 million (US\$19 million) since 1995 to research this poorly explored frontier.

As a result, a growing number of microbiologists are taking an interest in the ecology of the human gut. They are adapting tools previously developed for the study of microbes in oceans and soil to answer a range of questions. What lives in our gut? Do some natural gut microbes predispose us to diseases such as colon cancer? And can we change the make-up of our intestinal residents to improve our health? Gibson has even built a collection of artificial guts to study our internal microbial ecology under controlled laboratory conditions (see 'Roboguts', opposite).

Hidden world

The average human intestine contains about 1.2 kilograms of bacteria plus a smattering of yeasts. So far, few of these microbes have been characterized or even identified. But this dearth of information hasn't kept companies from promoting the health value of probiotics, which contain living bacteria, and prebiotics — nutrients designed to boost populations of beneficial bacteria already living in the gut.

Probiotic dietary supplements are available in just about any form imaginable, from tubes of liquid to capsules. Some yoghurts



Hard to swallow? Each human gut contains about 100 species of microbe.

and fermented milk drinks also promote their living contents. A typical online shop claims that its probiotic products can "strengthen the immune system, reverse the negative effects on the digestive tract of infections, antibiotics, alcohol ... treat symptoms of irritable bowel disease" and more. Worldwide, the pro- and prebiotics market is now worth about US\$6 billion.

But so far, the science behind these commercial boasts is rather limited. "There are a lot of bogus claims and vested interests," says Michael Blaut, head of gastrointestinal microbiology at the German Institute of Human Nutrition in Potsdam, and one of the researchers who helped to convince the EU to fund probiotics research. Although some clinical trials of probiotics have suggested a benefit, Gibson adds, few of these have been sufficiently rigorous. And even when probiotics

seem to work, he says, we know too little about the normal gut ecosystem to understand why.

Soon after it was established, the EU-funded network, which includes scientists from 16 countries, discovered that the gut ecosystem is much more diverse than previously thought. Microbiologists knew that their traditional techniques of isolating and cultivating individual microorganisms were not pulling out all of the species that we live with. Many gut bacteria are notoriously difficult to grow in culture — largely because they depend on the presence of other bacterial species. But few scientists had anticipated just how diverse the ecosystem would turn out to be.

To begin to quantify the diversity, gut researchers borrowed a method from soil and ocean microbiologists that relies on comparisons of the gene for a portion of the ribosome — the cellular machine that manufactures new proteins — known as 16S. The ribosome is so fundamental to the workings of the cell that it has changed little during evolution. That makes it easy to extract the 16S genes from all microorganisms in a single faecal sample using the DNA-amplifying polymerase chain reaction.

By looking for subtle differences between the sequences of these genes, microbiologists can gauge the number of different species present in the sample.

One surprise was that no two people have quite the same complement of bacteria. In unpublished work, molecular biologist Joël Doré of the INRA, the French agricultural research agency in Jouy-en-Josas, near Paris, has so far analysed the faeces of more than a dozen healthy adults and found the contents of each to be quite different. Although thousands of microbes can live in the gut, each person has only about 100 different species. "There is remarkably little overlap in the gut bacterial species between individuals," he says.

This is partly because of the haphazard way in which the bacteria arrive. Our guts start off in the womb completely sterile, but they are rapidly colonized with vaginal and

faecal bacteria during birth. Microorganisms from food and other environmental sources contribute to the mix during the first months of life. By the age of two at the latest, the average human gut hosts its full complement of microbial species, mixed and matched from a dozen or so dominant groups of bacteria and a longer list of rarer bacteria and yeasts.

From this point on, little changes until old age — a person's microbial complement seems to remain stable throughout adulthood, Doré says. But he has found that the faeces of people over 60 contain a much larger number of different bacteria than younger people. He suspects that the weakening barrier to new species may help explain why the elderly are more susceptible to gut infections and certain forms of cancer.

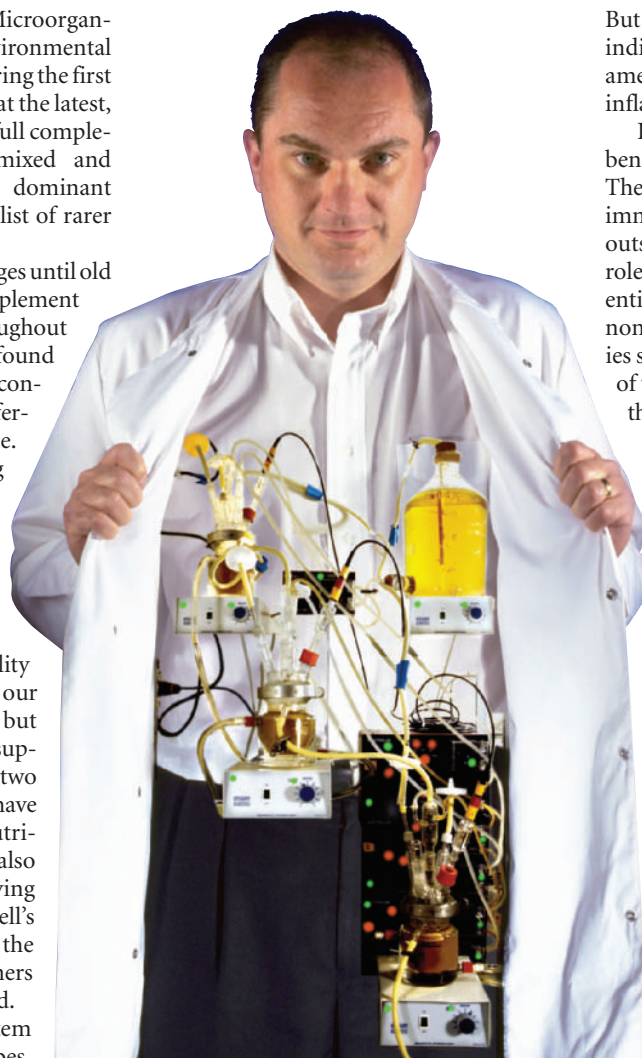
Colonic closed shop

The basis of the microbial stability that persists throughout most of our lives is still poorly understood, but is probably related to nutrient supply. By the time an infant is two years old, resident bacteria have monopolized every source of nutrients in the gut. They have also become interdependent, supplying nutrients to each other — one cell's waste is another's food. With all the nutrients accounted for, newcomers may find it hard to gain a toehold.

The stability of this ecosystem benefits not only the gut microbes, but also the human host. It prevents pathogenic bacteria, such as the various species of *Salmonella* that cause food poisoning, from taking up long-term residence. But it also means that probiotics cannot permanently change the make-up of the gut — they must be taken daily to have any effect.

What are the possible benefits? Individual species of bacteria are informally classified on a sliding scale of 'goodness' and 'badness'. Collectively, gut bacteria aid digestion by breaking down tough fibres, enzymes and other proteins. In addition, 'good' bacteria, such as species of *Lactobacillus*, *Bifidobacterium* and *Eubacterium*, are involved in fermentation reactions that produce organic acids that can be absorbed into the body and used as an energy source. 'Bad' bacteria, such as some members of the genus *Clostridium*, generate as by-products compounds including nitrosamines and cresols, which are possible carcinogens.

Commercial probiotic strains are, of course, 'good' bacteria. The probiotic milk products, yoghurts and capsules on the market generally contain *Lactobacillus* and *Bifidobacterium*. Most studies of their efficacy have been poorly controlled and have produced contradictory results, says Gibson.



Roboguts

The jumble of jars and tubes bubbling away in Glenn Gibson's laboratory smell pretty much as you would expect. They are models of the human gut. And Gibson (above), a gastroenterologist at the University of Reading, UK, uses them to investigate, among other things, the effects of new probiotics on gut microbial ecology.

He has infant guts, adult guts, ailing guts and more — around 20 in all. Each jar is a different section of colon seeded with the appropriate mix of microorganisms. Each assembly has a 'mouth' into which Gibson feeds meals, nutrients, probiotics or even harmful microbes.

The many tubes and ports give Gibson precise control over what goes on inside each of his models. For instance, by altering the speed at which food travels through them, he can give his artificial guts diarrhoea or constipation.

It may not be pretty, but Gibson says that the models are good enough to give him a fairly clear view of the events that would otherwise be hidden in the murky depths of our entrails.

But a handful of well-designed clinical trials indicates that some such bacteria may help ameliorate diarrhoea¹⁻³ and some types of inflammatory bowel disease^{4,5}.

Less well documented are the claims for beneficial stimulation of the immune system. The gut, with its massive blood supply, is the immune system's primary contact with the outside world, and gut bacteria seem to play a role in teaching the immune system to differentiate between dangerous invaders and non-hostile challenges. Although some studies suggest that probiotics can affect features of the immune system, few have shown that these changes are beneficial to health.

A notable exception is a long-term study supported by the Finnish Academy of Sciences, in which pregnant women from families prone to allergies ate *Lactobacillus rhamnosus* daily. After delivery, the bacteria were given daily to the babies for the first six months of their lives. The treated infants were much less prone to allergic reactions such as eczema than controls who did not get the bacteria^{6,7}. Erika Isolauri, an immunologist at the University of Turku who led the study, is now trying to determine how the treatment works. She suspects that the probiotics shift the balance between pro- and anti-inflammatory factors in the developing gut.

Friend or foe?

Other studies to assess the health benefits of probiotics are under way. With funding from the EU, for example, Doré is setting out to test a combination of *Bifidobacterium animalis* and a type of prebiotic sugar known as FOS on the gut ecosystems of young and old people in France, Germany, Sweden and Italy. "We are testing faecal samples to see whether the level of *Bifidobacterium* really does rise with this treatment, as would be expected," he says. His team will also assess how the levels of toxic and potentially carcinogenic compounds in the gut rise and fall with treatment by exposing cell cultures to extracts from the subjects' faeces. The researchers hope to determine whether suppressing 'bad' bacteria with pre- and probiotics might protect against colon cancer.

Francisco Guarner, a gastroenterologist at the Vall d'Hebron Hospital in Barcelona, Spain, is helping to organize an EU-backed clinical study involving 360 patients chronically suffering from one of two types of inflammatory bowel disease — ulcerative colitis or Crohn's disease — at centres in Ireland, Spain, Finland and France. The patients, all in remission, receive either *Lactobacillus salivarius* or *Bifidobacterium infantis*, two species that reduce gut inflammation in lab animals. The researchers then test the patients' saliva for marker molecules associated with inflammation. "Animal studies

show that inflammatory disorders of the bowel may be helped by making the gut microbes less aggressive," says Guarner. He hopes that the probiotics will extend the patients' remission so they can reduce their reliance on immunosuppressive drugs, which have severe side effects.

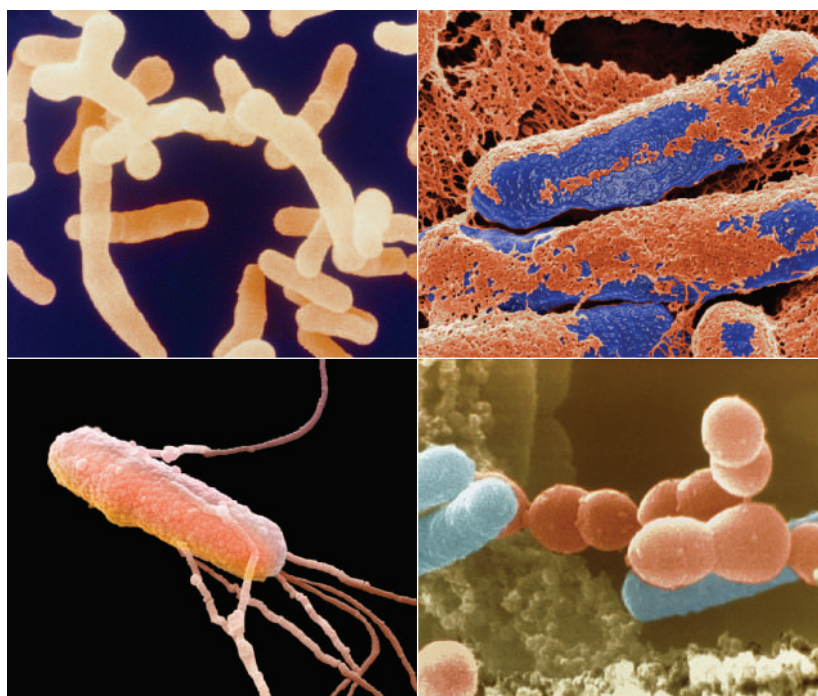
Outside the EU network, Gibson is running trials at six British centres using *Lactobacillus plantarum* together with a second type of prebiotic sugar called GOS in various kinds of inflammatory bowel disease. Gibson's hypothesis is that the yeast *Candida* causes the symptoms, and he hopes that the probiotics will outcompete its growth. In a separate study, he is testing his hypothesis that certain 'bad' bacteria contribute to ulcerative colitis by generating toxic sulphur compounds such as hydrogen sulphide, which smells of rotten eggs. He is giving patients FOS and GOS to stimulate the growth of competing 'good' bacteria to see whether this eases the symptoms.

Microbes on trial

In the next few years, these and other studies will help to determine how beneficial probiotics actually are. In the meantime, other salient questions are being addressed. Can we, for example, assume that probiotics are safe? One potential problem is that many probiotic strains have genes that allow them to resist antibiotics, which they might pass on to pathogenic bacteria. To address these concerns, Herman Goossens of the University of Antwerp in Belgium has acquired more than 200 commercial probiotic strains — the world's largest collection. He is systematically analysing them for their potential to transfer antibiotic resistance genes, and is also testing for any direct toxic effects that they may have.

Some experts believe that another important step will be to read the genomic sequence of every species of microbe that can colonize the human gut. They argue that a complete genomic databank would make it much easier to select species for specific probiotic effects.

To that end, the Defense Advanced Research Projects Agency, a research arm of the US military, is sponsoring a project to read all the genomes in the gut ecosystem with the same 'shotgun' method used for the privately funded effort to sequence the



'Good' bacteria such as *Bifidobacterium* (top left) and *Lactobacillus* (blue, bottom right) may help to ward off pathogens such as *Salmonella* (bottom left) and *Clostridium botulinum*.

human genome. This approach avoids the need to separate out individual organisms. Instead, fragments of all the genomes are read off together. Computer algorithms then reassemble the fragments on the basis of overlapping sequences into complete genomes. "It's possible to conceive of doing this because the cost of sequencing has come right down," says Claire Fraser, director of The Institute for Genomic Research in Rockville, Maryland, where the work will be done.

Even if probiotics and prebiotics prove to have only modest health benefits, some scientists are considering the possibility of souping them up with genetic engineering. Many proponents of probiotics reject this idea, saying that it would be too hard to convince the public to eat live, genetically modified bacteria. But among the traits that would be useful to engineer are the ability to survive the acid environment of the stomach, a bit of 'stickiness' to help bacteria adhere to the gut lining, and so take residence for longer, and the ability to produce organic acids. Bacteria might even be engineered to deliver drugs, vitamins or vaccines^{8,9}.

There is already evidence that some bacteria can serve as efficient delivery vehicles. For example, Lothar Steidler of Ghent University in Belgium and his colleagues have shown that *Lactococcus lactis* genetically modified to secrete the anti-inflammatory molecule interleukin-10 can reduce colitis in mice¹⁰. A version for humans has also been developed

that includes safety features to prevent the escape of the inserted gene into the environment¹¹. A small clinical trial of this microbe is planned in Amsterdam, marking the first use of a genetically engineered bacterium as a therapeutic agent.

Unfortunately for the scientists studying probiotics, the only way forward is to delve into more human waste. Doré says he recently felt a pang of regret over that fact on a trip to visit some oceanographer friends in Marseille. "I looked out into the Mediterranean and thought: 'I'm obviously working on the wrong ecosystem,'" he sighs. But the scientific challenges presented by gut bacteria are interesting enough to keep him going, he says. "It makes up for the unpleasantness."

Alison Abbott is *Nature's* senior European correspondent.

1. Guandalini, S. *et al.* *J. Pediatr. Gastroenterol. Nutr.* **30**, 54–60 (2000).
2. D'Souza, A. L., Rajkumar, C., Cooke, J. & Bulpitt, C. J. *Br. Med. J.* **324**, 1361–1366 (2002).
3. Cremonini, F. *et al.* *Aliment. Pharmacol. Ther.* **16**, 1461–1467 (2002).
4. Rembacken, B. J., Snelling, A. M., Hawkey, P. M., Chalmers, D. M. & Axon, A. T. R. *Lancet* **354**, 635–639 (1999).
5. Gionchetti, P. *et al.* *Gastroenterology* **119**, 305–309 (2000).
6. Kalliomäki, M. *et al.* *Lancet* **357**, 1076–1079 (2001).
7. Kalliomäki, M., Salminen, S., Poussa, T., Arvilommi, H. & Isolauri, E. *Lancet* **361**, 1869–1871 (2003).
8. Seegers, J. F. M. L. *Trends Biotechnol.* **20**, 508–515 (2002).
9. Wood, B. J. B. & Warner, P. J. (eds) in *The Lactic Acid Bacteria* Vol. 3, 261–290 (Kluwer Academic, New York, 2003).
10. Steidler, L. *et al.* *Science* **289**, 1352–1355 (2000).
11. Steidler, L. *et al.* *Nature Biotechnol.* **21**, 785–789 (2003).



Milton and Galileo would back *BMJ* on free speech

Arguments, crazy ideas and open communication are the lifeblood of science.

Sir—Your News story “Medical journal under attack as dissenters seize AIDS platform” (*Nature* **426**, 215; 2003) was a fair report of researchers’ objections to rapid responses being posted on the website of the *British Medical Journal* (*BMJ*) by people who are sceptical about a link between AIDS and HIV. As editor of the *BMJ*, however, I find it disturbing to see scientists arguing for restrictions on free speech. Surely open communication and argument is a fundamental value of science?

John Milton put the argument better than anybody in 1643, in his pamphlet *Areopagitica*. “Give me,” he wrote, “the liberty to know, to utter, and to argue freely according to conscience, above all liberties. ... [W]ho ever knew Truth put to the worse, in a free and open encounter? ... Yet is it not impossible that she [truth] may have more shapes than one ... [I]f it come to prohibiting, there is not aught more likely to be prohibited than truth itself; whose first appearance to our eyes, bleared and dimmed with prejudice and custom, is more unsightly and unpalatable than many

errors ... Where there is much desire to learn there of necessity will be much arguing, much writing, many opinions; for opinion in good men is but knowledge in the making.”

We should never forget Galileo being put before the inquisition. It would be even worse if we allowed scientific orthodoxy to become the inquisition.

I’m not arguing that those who doubt the link between HIV and AIDS are right, but I want to keep our threshold for posting rapid responses as low as possible.

How, I’m legitimately asked, does this fit with an editorial code I have drafted saying: “Editors should take all reasonable steps to ensure the accuracy of the material they publish.” My first reaction is that perhaps “accuracy” is the wrong word to use. As editors we receive thousands of manuscripts containing millions of assertions. We can’t possibly check every “fact”, and distinguishing fact from opinion is not as straightforward as it sounds.

The answer, I think, lies in transparency. Our rapid responses are clearly unfettered

debate full of crazy ideas, false logic, and unreadable, mis-spelt prose as well as some literary and scientific gems. What you see is what you get. In contrast, original articles have been as rigorously peer-reviewed as we can manage, with the recognition that peer review itself is a deeply flawed process.

Your News story states: “The dispute crystallizes the conflict in the Internet era between a journal’s desire to experiment with open electronic debate, and its fundamental obligation to its readers to provide them with authentic information.” I don’t agree that there is a conflict. The beauty of the electronic world is that we can have no-holds-barred debate alongside greater selectivity. On our website you can do a search that includes or excludes rapid responses. I suggest that those who want to see the world as it is — rather than how they would like it to be — include rapid responses in their search.

Richard Smith

*BMA House, Tavistock Square,
London WC1H 9JR, UK*

X-ray clues to viability of loop quantum gravity

Sir—The unification of quantum mechanics with gravity is the most pressing question in theoretical physics today. However, experimental feedback to the theorists has been sorely lacking. Astrophysicists are now beginning to probe the behaviour of gravity at quantum (microscopic) scales.

For example, Igor G. Mitrofanov (*Nature* **426**, 139; 2003) described a possible constraint on a leading theory, loop quantum gravity, based on the polarization of high-energy radiation from astrophysical sources. The high-energy photons have to travel cosmological distances to reach us, allowing small effects of quantum gravity to reveal themselves. This specific constraint depends on the reported detection of polarization from a γ -ray burst, which has yet to be confirmed, hence Mitrofanov cautioned readers to await confirmation of this measurement before concluding that loop quantum gravity is not viable.

There is no need to wait. The constraint on the polarization of γ -rays applies equally to the polarization of X-rays, for which there are 30-year-old measurements. The X-ray polarization

of the Crab nebula, a thousand-year-old remnant of an exploded star, was first measured by Novick and collaborators¹ in 1972 and confirmed by a different instrument four years later². The observed X-ray polarization from the Crab nebula is in strong conflict ($\chi < 10^4$) with the predictions of loop quantum gravity, if the effects of quantum gravity depend linearly on photon energy.

Philip Kaaret

*Harvard-Smithsonian Center for Astrophysics, 60
Garden St, Cambridge, Massachusetts 02138, USA*

1. Novick, R. et al. *Astrophys. J.* **174**, L1–L8 (1972).

2. Weisskopf, M. C. et al. *Astrophys. J.* **208**, L125–L128 (1976).

Return of bone archives is a loss to humanity

Sir—As a member of the panel that published the Palmer report supporting the principle of repatriation of ancestral remains (“Bone archives face prospect of dispersal” *Nature* **426**, 109; 2003), but also as the sole author of a statement of dissent, I argue that competing concerns must be weighed carefully on a case-by-case basis.

There must be a balance between the concerns of claimant communities on the one hand and, on the other, the loss to humanity resulting from wholesale return on request. It is clear to me that this

balance will be destroyed if the return of human remains on request becomes the order of the day.

There is a strong case for the retention of human remains in the scientific collections of museums and university departments. Research into collections at the Natural History Museum in London over recent years has enabled surgeons to explore improved surgical techniques on damaged knee joints. It has increased our understanding of diverse human groups’ responses to malaria, tuberculosis and other diseases, and has trained forensic anthropologists in methods which they then used to help identify bodies found in Bosnia’s mass war graves.

The benefits that these collections provide to humanity are not sufficiently recognized in the Palmer report. Its recommendations that there should be return on request from claimant communities or individuals within those communities are tantamount to mandatory repatriation.

Neil Chalmers

*The Natural History Museum, Cromwell Road,
London SW7 5BD, UK*

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

The hole truth

What's news (and what's not) about the ozone hole.

Susan Solomon

The discovery in 1985 of an unprecedented 'hole' in Antarctica's ozone layer heralded the beginning of one of the most influential environmental stories of the late twentieth century¹. The tale of the missing ozone at the bottom of the world was made all the more intriguing by its strong seasonal behaviour. Ozone levels decline rapidly each year in August and September (Antarctic spring), and the hole is typically at its deepest by late September to early October. It then largely fills in through mixing of surrounding ozone-rich air by late January, ready for the next year's cycle.

Each Antarctic spring in the late 1980s and early 1990s, enthusiastic scientists shared the news about the hole with a fascinated and equally enthusiastic public. Slowly, the mystery surrounding events in the Great White South evolved into understanding for both groups². The root cause of the hole was identified as a number of industrially produced chemicals. Policy-makers worldwide quickly agreed on the Montreal Protocol to phase out these chemicals, and by the late 1990s global production of these gases had dropped by more than 90% (see 'How can we tell if the Montreal Protocol is working?', overleaf). Like many scientists involved in this sea-change in scientific and public knowledge, I have been heartened by widespread interest in the issues, and encouraged by the ability of even school-children to understand the key elements of the story.

The ozone hole has been considered by many to be the success story of global environmental policy ever since the Montreal Protocol came into force. But in the past few years I have grown concerned about a public that seems to be more confused than intrigued by the news reports that greet the appearance of the hole each year. Despite the apparent success of the Montreal Protocol, the elimination of the ozone hole may seem to progress at a remarkably uneven rate. And as the hole changes in size and shape (Fig. 1, overleaf), communicating to the public what this means has become much harder.

I feel that this presents the scientific community with a new challenge. Given that global production of ozone-damaging compounds is now nearing zero, yet ozone depletion will continue for many years to come, how can scientists help to preserve the remarkable success in public understanding of the ozone hole? How can we best explain why this is so in language understandable to



The advent of a hole in the ozone layer over Antarctica brought the continent into the media spotlight.

students, teachers, scientific colleagues in other fields — indeed to everyone who shares our interest in a phenomenon that is not just scientific but also historic, not just technical but also sociological? How do we distinguish between what is news and what is not as the ozone hole moves into an expected period of very slow recovery? Are there any observable connections between the behaviour of the ozone hole and whether or not the Montreal Protocol is working? If not, how can we tell if the protocol is on or off track?

Cause and effect

Within five years of the ozone hole's discovery, the cause had been established. Direct observations of chemicals in the atmosphere showed that an increase in chlorine concentrations in the stratosphere was the key agent responsible³. This rise was mainly due to chlorofluorocarbons (CFCs) — long-lived chemicals produced by the chemical industry and used variously as, for example, coolants in refrigerators and air conditioners, foam-blowing agents and solvents.

The chemical reactions that destroy ozone can occur with unparalleled efficiency in Antarctica because of the very cold conditions in the polar stratosphere during the winter and spring months. These allow special high-altitude clouds to form as soon as temperatures dip below -85°C (ref. 4). Cold surfaces within these clouds serve as the sites for very rapid reactions that convert

inactive chlorine into its ozone-destroying forms. The hole appears during the Antarctic spring because the key ozone-destroying reactions are initiated by sunlight, a factor largely absent during winter months.

Why is there no Arctic hole? Ultracold temperatures are far more widespread and persistent in the Antarctic winter and spring atmosphere than in the Arctic, where the flow of air over the Himalayas and Rocky Mountains and land-sea temperature contrasts can generate very large 'atmospheric waves'. On the ground, we experience many of these waves as the passage of storms, and some travel upwards to the stratosphere, ultimately mixing warmer mid-latitude air with cold polar air. So the varied topography of the Northern Hemisphere gives it a greater number of atmospheric waves and a warmer polar stratosphere in the winter and spring on average than in the south⁵.

In brief, the ozone hole is driven by a blend of three factors: excess chlorine, which is why the ozone hole is a recent phenomenon; cold temperatures, which account for its occurrence in Antarctica; and sunlight, which explains why the hole opens up in spring, as light returns to the polar cap.

Repairing the damage

As production of CFCs ceases, how long will it take for the reactions fuelling ozone loss to subside? The key is the atmospheric lifetime of the CFCs. Many air pollution problems, such as acid rain, are linked to

chemicals that have short lifetimes in the atmosphere because they dissolve in rainwater within a few days, or are rapidly destroyed by chemical reactions. Likewise, most of the chlorine from natural sources, such as sea spray or volcanoes, dissolves quickly and effectively in rainwater, preventing much of it from reaching the stratosphere⁶. But the major CFCs have atmospheric lifetimes of 50–100 years, allowing them not only to reach the stratosphere in significant quantities but also to remain there long after emissions have stopped (see ‘How can we tell if the Montreal Protocol is working?’, opposite).

By 2001, global reporting of the production and emission of CFCs revealed a dramatic drop from peak values in the late 1980s and this trend was confirmed by air sampling studies^{7–12}. Nonetheless, the long lifetimes of the megatonnes of CFCs already in the atmosphere mean that clear and lasting improvements in the ozone hole can take place only slowly. Removal of the ozone-destroying chlorine and hence return of the Antarctic ozone layer to its natural state will take decades.

Variable pattern

But even in the relatively inactive Southern Hemisphere, with its less variable topography compared with the north, atmospheric waves and weather vary from one winter or spring to another. Just as a particularly warm spring can occur in, say, Wellington, New Zealand, one can also occur in the Antarctic stratosphere, resulting in less ozone depletion for that year. Satellite data confirm the distorting effects of atmospheric waves on the extent of the ozone hole from year to year (see Fig 1)⁵.

It is the annual reporting of these ups and downs in the ozone hole’s area (and occa-

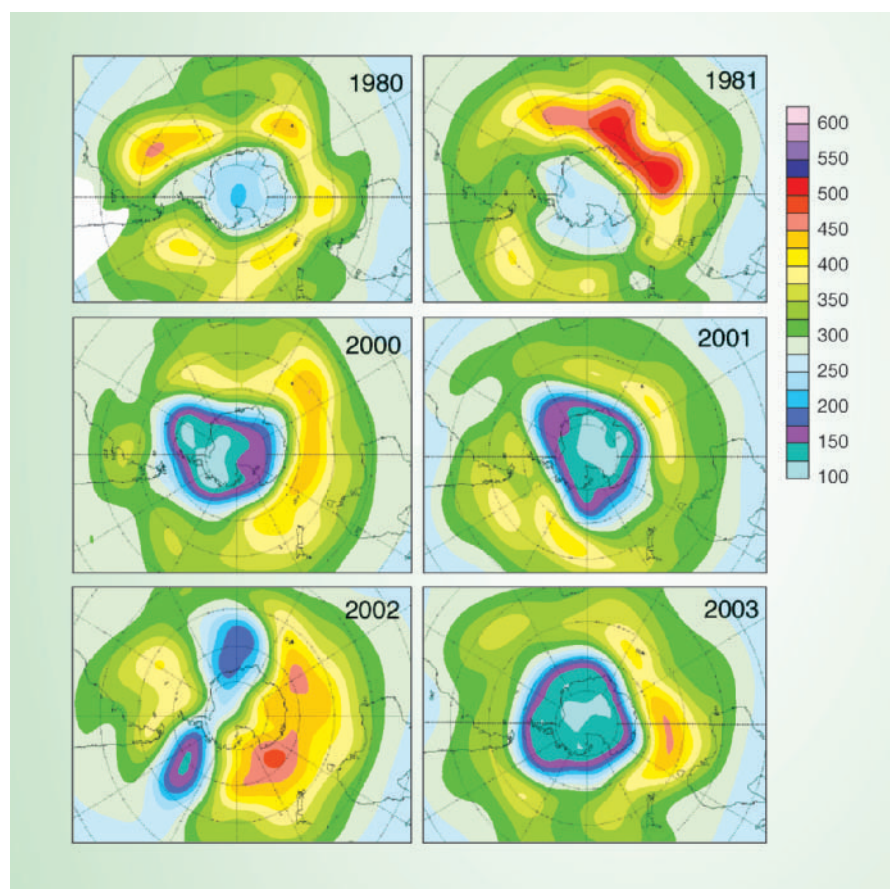


Figure 1 Satellite maps of total ozone over Antarctica on 24 September, when the ozone hole is near its annual peak, in 1980, 1981, 2000, 2001, 2002 and 2003. The colour scale shows the amount of ozone in Dobson units, indicating the depth of the hole. The images are based on multiple satellite records and analyses and are available from many sources, including the World Ozone and Ultraviolet Radiation Data Centre in Toronto, Canada, <http://woudc.ec.gc.ca/cgi-bin/selectMap>.

sionally depth) that seems to be causing general confusion. For example, the ozone hole in 2000 was large, leading much of the media — including respected institutions such as the BBC¹³ — to speculate that recovery might take longer than had been expected. Other news sources suggested that it was a sign of a reversal in recovery, or a problem with the Montreal Protocol.

In contrast, the 2002 ozone hole was much smaller and split in two, an unprecedented event. A massive Antarctic wave led to stratospheric temperatures in September that were higher than observed in the previous 50 years, affecting both the area and depth of the hole¹¹. Whether this was a rare but random event dictated by the intrinsic chaos of atmospheric dynamics (as for the ‘storm of the century’) or the beginning of a more systematic shift is currently unknown. It certainly did not signal that the protocol was working even faster than expected, although numerous reports that I saw on local and national television suggested exactly that. It was also interpreted that way by some industrial representatives according to a report in the *New York Times*¹⁴.

For us to know whether the Montreal Protocol’s future will match its past success will require continued observations of global changes in CFC levels (see ‘How can we tell if the Montreal Protocol is working?’, opposite). These measurements should therefore be highlighted whenever the protocol’s status is being discussed by scientists or journalists. In contrast, linking fluctuations in hole size with the protocol’s effectiveness should be discouraged. Public understanding is likely to be best served if scientists stress that year-to-year changes in ozone-hole area (or depth) are more likely to reflect changes in stratospheric ‘weather’ conditions that influence ozone loss over the short term, rather than long-term issues such as whether the protocol is working or not.

Future scientific questions that might be linked to the fate of the ozone hole include whether global warming will begin to affect ozone depletion (either through effects on waves or on temperature); whether the Arctic stratosphere will change to be more like the Antarctic in coming centuries; and how possible changes in the atmospheric concentration of methane or nitrous oxide might affect ozone-hole chemistry. Insights on



Monitoring the levels of CFCs in air samples is key to assessing the success of mitigation efforts.

H. MORGAN/SPL

these issues obtained by monitoring the ozone hole will surely be of interest to the public, but will not arise from observations from a single year, or even over a few years, underscoring the need for care when discussing them.

Are annual reports on the development and size of the ozone hole still newsworthy in view of these considerations? One might imagine that the time has passed for annual press statements on a phenomenon that is so well known and understood, but like many issues in science, this one has a range of devoted and rightfully interested parties. People throughout the Southern Hemisphere reasonably want to know as early as possible what is happening above Antarctica, because severely ozone-depleted air can occasionally be transported to populated areas of South America, New Zealand and even as far north as Australia. But the ozone hole belongs not just to the people of the Southern Hemisphere nor to interested scientific audiences, but to a wider world whose curiosity has been aroused. In their annual announcements and reports, both scientists and journalists will best serve the public by taking that into account.

Because there are several sources of information on the ozone hole, there has been a rush in previous years to be the first to announce the appearance of the hole. Taking more time on the part of scientists to make clear just what the story is and to ensure that the analyses are complete would foster greater public understanding. Taking more care on the part of journalists to avoid leaps to unfounded policy inferences would bolster the probity that is the hallmark of quality science reporting. Scientists and journalists will be best placed to communicate with each other and with the public if together they ensure that any routine announcements on the ozone hole not only describe what is occurring each year but also what that does and does not mean.

Susan Solomon is at NOAA Aeronomy Laboratory, Boulder, Colorado 80305, USA.

- Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207–210 (1985).
- Cagin, S. & Dray, P. *Between Earth and Sky* (Pantheon Books, New York, 1993).
- Scientific Assessment of Ozone Depletion: 1991* (Rep. 25, World Meteorological Organization, Geneva, 1991).
- Solomon, S., Garcia, R. R., Rowland, F. S. & Wuebbles, D. J. *Nature* **321**, 755–758 (1986).
- Scientific Assessment of Ozone Depletion: 1998* (Rep. 44, World Meteorological Organization, Geneva, 1998).
- Tabazadeh, A. & Turco, R. P. *Science* **260**, 1082–1086 (1993).
- McCulloch, A., Ashford, P. & Midgley, P. M. *Atmos. Environ.* **35**, 4387–4397 (2001).
- Montzka, S. A. *et al.* *Nature* **398**, 690–694 (1999).
- Walker, S. J., Weiss, R. F. & Salameh, P. K. *J. Geophys. Res.* **105**, 14285–14296 (2000).
- Prinn, R. G. *et al.* *J. Geophys. Res.* **105**, 17751–17792 (2000).
- Allen, D. R., Bevilacqua, R. M., Nedoluha, G. E., Randall, C. E. & Manney, G. L. *Geophys. Res. Lett.* **30**, 1599–1602 (2003).
- http://www.afeas.org/prodsales_download.html
- <http://news.bbc.co.uk/1/hi/sci/tech/916037.stm>
- Revkin, A. C. U.S. to Seek Support for Ozone Exemptions at Meetings. *New York Times* (10 November 2003).

How can we tell if the Montreal Protocol is working?

The ozone hole grew during the 1980s as emissions of chlorofluorocarbons and other gases built up in Earth's atmosphere and ozone-depleting chemistry took place at ever-faster rates. By the late 1990s, enough nations had signed the Montreal Protocol to drastically reduce global emissions of most of the key gases. How can we be confident that these actions are working?

Monitoring these gases in our atmosphere provides direct proof. The great bulk of industrial chemicals that contribute to changes in the ozone layer are remarkably inert and so survive in the atmosphere for decades or even centuries. The mix of factors that controls the build-up and decline of such gases — production, emission and lifetime — can be illustrated by data for one major chlorofluorocarbon, CFC-11 (see inset).

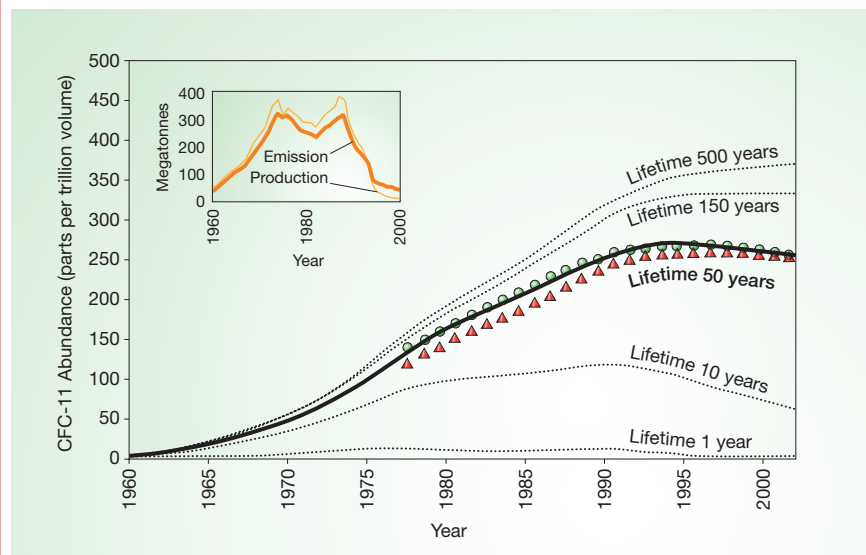
Global production and emission for CFC-11 (see inset and refs 7, 12) are based upon industrial estimates, and are thought to be accurate to about 15%. Production is not necessarily the same as emission in a given year because of delayed release from existing materials, such as thermal foam insulation in older buildings or refrigerators⁷. Consumers and policy-makers who wish to accelerate the recovery of the ozone should consider responsible recovery of such materials, as is done when the materials in discarded refrigerators are recycled rather than shredded.

The third factor, lifetime, is reflected in the comparison of the observed and calculated abundances of CFC-11. Observations from

both hemispheres^{8–10} (data points) are shown alongside the calculated abundances based on global emission estimates (curves) and adopting atmospheric lifetimes of 1, 10, 50, 150 or 500 years. These show that the atmospheric lifetime of CFC-11 is about 50 years. If it were much shorter, the gas's abundance could never have built up to its present value and would have dropped more steeply by now; much longer and the concentrations would have peaked at higher values and would be continuing to increase slowly today despite the greatly reduced emissions of recent years.

The slow decline in abundances now being observed by scientists worldwide supports, within data uncertainties, the true 50-year lifetime of CFC-11 and illustrates the success of the Montreal Protocol. More evidence that the protocol is working comes from the gradual closing of the gap between Northern and Southern Hemisphere abundances measured over the past decade. During the build-up of emissions in the 1970s and 1980s there were far larger releases of CFC-11 in the more heavily populated Northern Hemisphere, but as emissions declined after 1990, atmospheric mixing acted to even out the global distribution. Similar behaviour has been documented for many other chlorofluorocarbons. These data show that the Montreal Protocol is working, while emphasizing the need for patience as we seek signs of genuine recovery in the ozone layer over future decades.

S.S.



CFC-11 in the atmosphere. The curves represent predicted abundance of CFC-11 for different atmospheric lifetimes. Actual data for Northern (circles) and Southern (triangles) Hemispheres show that CFC-11's atmospheric lifetime is about 50 years. Data points compiled from refs 8 and 9; data from ref. 8 have been adjusted by 2% to align calibrations within uncertainties; curves adjusted upward by 10% within uncertainties.

The curious naturalist

Niko Tinbergen was one of the founders of ethology.

Niko's Nature

by Hans Kruuk

Oxford University Press; 2003. 406 pp. £20

John Krebs

It was 1965. Niko Tinbergen, a small, white-haired, bespectacled man of modest manner, was in full flow in one of his enchanting undergraduate lectures. Suddenly, he leapt onto the front bench of the steeply raked Victorian lecture theatre, strode from bench to bench past astonished students right up to the back, snatched a newspaper from the grasp of an inattentive undergraduate, and pranced gazelle-like down the benches to the front. And all without pausing in his explanation of how some oystercatchers pierce mussel shells by stabbing whereas others hammer them open. I remember being both relieved and astounded: relieved that I wasn't the one with the newspaper, and astounded by Tinbergen's athletic burst. As a second-year undergraduate at the University of Oxford, I didn't know that as a young man he had played international hockey in Holland.

His athletic prowess is one of many charming glimpses of Tinbergen in Hans Kruuk's thoroughly researched biography, an account that is affectionate and sympathetic, while at the same time being dispassionate and appraising. Kruuk was a student (or "pupil" as Tinbergen called them), friend and admirer of Tinbergen and so writes with the authority of first-hand knowledge. But this is no dewy-eyed romantic reinvention of a hero 20 years after his death.

Tinbergen was a major figure in twentieth-century biology. As co-founder of the biological study of animal behaviour known as ethology, he shared the Nobel Prize with Konrad Lorenz and Karl von Frisch in 1973 and, like all creative scientists, he set the research agenda for generations of followers.

His approach to the science of animal behaviour is perhaps best captured in the title of one of his classic books, *Curious Naturalists*. Tinbergen observed and wondered. He turned observation into ingenious experiment, description into analysis, and natural history into science.

Growing up in the Netherlands in the early twentieth century as the son of a school teacher, he joined the Dutch Youth League for Nature Study. This helped him develop his skills as a naturalist and his talent as an artist — the biography is illustrated with many of his sketches. Perhaps because the Netherlands is a small, crowded country, Tinbergen's love of nature was expressed in his eye for detail and the beauty in the

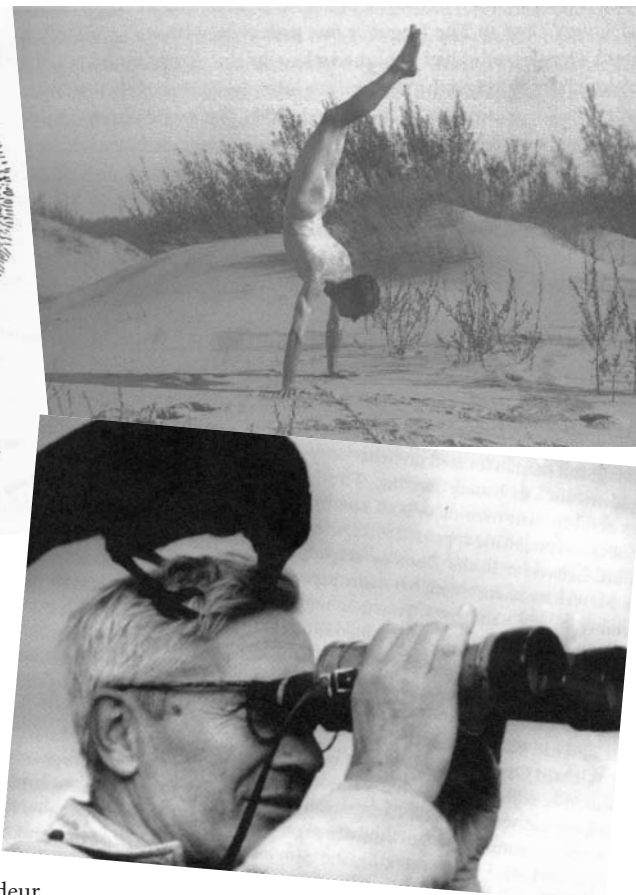


Back to nature? The athletic Niko Tinbergen (photos) spent much of his time doing fieldwork, observing birds and making sketches of them.

apparently commonplace — such as gulls on a sand dune and sticklebacks in a ditch — rather than in the grandeur of great tracts of wilderness.

Looking back at Tinbergen's scientific work, Kruuk sees some contradictions. Tinbergen created and cultivated the idea that great insights into the natural behaviour of animals could be derived from simple, elegant experiments. But, as Kruuk skilfully exposes, many of Tinbergen's experiments were far from perfect in their design, execution and analysis. This did not diminish their importance and impact either in the scientific community or in the eyes of the general public, however. Tinbergen was a great popularizer of scientific research, and many wildlife programmes on television took as their source the work of Tinbergen and his students. This reached its zenith when his film *Signals for Survival* won the Italia Prize for TV documentaries in 1969.

In his early career at the University of Leiden, Tinbergen created, together with Lorenz, a theoretical framework for ethology, summarized in his seminal 1951 book *The Study of Instinct*. Within a few years of its publication, this framework of innate drives, releasing stimuli and consummatory behaviours had come under telling attack by Danny Lehrman and Robert Hinde. Nevertheless, the clarity of Tinbergen's argument,



even when it was later shown to be inadequate, shone a light for others to see a way forward.

By contrast, much of Tinbergen's work following his move to Oxford in 1949 was less influenced by theory, comprising instead a series of elegant stories that were increasingly concerned with the function as well the causation of behaviour, each one a gem but not knitted together as a coherent whole. This was despite the fact that with his new colleagues — notably David Lack, Charles Elton, George Varley, Arthur Cain and Bernard Kettlewell — he had the potential to weave together ethology, population biology and evolution. This blending, to create ethology's offspring behavioural ecology, had to wait for the next generation.

A recurrent theme in Kruuk's book is Tinbergen's guilt about being paid to do what he liked best as a hobby: watch animals in the wild. As he wrote in *Curious Naturalists*: "It is I think only natural for a man to have occasional doubts about the value of what he is doing, at any rates such doubts have often occurred to me... A little devil seems to look over our shoulder and to take great pleasure in kindling this spark of doubt."

This was a major motivator for him to

turn in the last part of his career to the ethology of man. In the 1960s, Desmond Morris's *The Naked Ape* and Lorenz's *On Aggression* — both engaging, if flawed, books — had thrown open the whole debate about the extent to which present-day human behaviour can be explained in terms of our evolutionary past. If our evolutionary history provides explanations, perhaps it also provides clues to the treatment of dysfunctional or abnormal human behaviour.

Tinbergen's venture into this area did not attract widespread support. His work, with his wife Lies, on the ethology of childhood autism was seen by many as too anecdotal and simplistic, and his Nobel lecture, in which he talked about both autism and the rather wacky Alexander technique for improving body posture, would, in the words of Kruuk "be best forgotten".

In addition to his comprehensive account of Tinbergen the scientist, Kruuk also discusses Tinbergen the man. He offers insights into Tinbergen's relationships with his wife and children, his siblings, including his older brother Jan, who won a Nobel Prize for economics in 1969, and with his students and colleagues. Kruuk also discusses in detail the recurrent depression that afflicted Tinbergen, especially during the last 20 years of his life, which impaired his capacity for work.

Kruuk succeeds in conveying the complexity of Tinbergen's personality. Tinbergen was self-effacing, modest and charming, totally without pomposity, haunted by self-doubt, but was at the same time demanding of himself and others, ambitious, charismatic and a natural showman and leader. He was legendary for his rigorous powers of analysis (all of us who participated in his weekly seminars have our own favourite stories of his clinical dissection of any hapless speaker with leaky intellectual plumbing), but he often relied on intuition and creative leaps in drawing conclusions.

What of his lasting contribution to science? Where has ethology gone since Tinbergen's death? Some of the people Kruuk quotes think it is "as alive and strong as ever", whereas others say it has evolved into behavioural ecology and animal cognition.

The prognosis offered some 20 years ago by Robert Hinde in his Niko Tinbergen lecture is more inclusive. He argued that the lasting influence of ethology, and therefore of Tinbergen's work, would be felt by the interaction of its ideas and approaches with many other fields, such as anthropology, neuroscience, psychiatry, psychology, evolution and ecology. It is now widely accepted that it is possible to analyse the behaviour of animals in their natural environment with rigour and precision. Niko showed us how to do it. ■

John Krebs is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Chemistry's bag of tricks

**From Elements to Atoms:
A History of Chemical Composition
(Transaction 92-4)**

Robert Siegfried
American Philosophical Society: 2003.
278 pp. \$24

Bernadette Bensaude-Vincent

As a scholar for more than 40 years, Robert Siegfried says he has witnessed the emergence of the history of science as an academic specialty, and would like to see it included in the scientific curriculum. With such didactic purposes in mind, he deliberately runs counter to the current historiographical mainstream.

Siegfried's view is that chemistry evolved from a metaphysical idea of composition into an experiential, positive idea of composition, forged by Antoine-Laurent Lavoisier and completed by John Dalton's atomic theory. In so doing, Siegfried advocates a positivistic view that has been heavily criticized by a generation of historians of science. He also purposefully writes history from a present-day perspective.

On top of that, he focuses on the chemists' philosophy of matter, even though the interest of many historians shifted long ago from concepts to practices and the social dimensions of science. He quotes several recent historical studies of the same period, but never engages in historiographical debates. This is "my story", he claims, just one interpretation among many others. In his view, history is "a bag of tricks that we play on the dead".

But despite his old-fashioned approach, Siegfried's account of eighteenth-century chemistry is not too dissimilar from other recent descriptions. Like Frederic L. Holmes in his book *Eighteenth-century Chemistry as an Integrative Enterprise* (University of California, 1989), he stresses the importance of salt chemistry, which provided the empirical core of modern chemistry. And like Ursula Klein (*Science in Context* 7, 163; 1994), he assumes that analysis and synthesis played a key role in the emergence

of an empirical concept of composition. He agrees with most historians of chemistry that combustion was not the major issue discussed in the eighteenth century, and that phlogiston theory was by no means the overarching doctrine that dominated all chemical works before Lavoisier. In fact, it only really became a theoretical framework when chemists started investigating dozens of 'airs' in the 1770s.

In other words, Siegfried does not read eighteenth-century chemistry 'backwards' as an empty period preparing the stage for Lavoisier's chemical revolution. Instead, he provides a fine and nuanced account of many of the achievements that pre-dated

Lavoisier, such as the measurement by Torbern Bergman of the relative quantities of phlogiston in metals.

Even so, Siegfried's choice of a present-day perspective distorts this penetrating and thoughtful insight into chemistry's past. How relevant is a historical narrative that focuses on the absence of what would emerge later? In this book, Siegfried repeatedly points out what was lacking — a clear distinction between concepts, and a clear notion of what chemical substances are — and traces anticipations and pre-figurations of notions that would only be

made conscious and explicit by Lavoisier or Dalton. Wouldn't it be more useful for students to try to understand how chemists of the past constructed their own views and notions? But Siegfried assumes that chemists before Lavoisier had no theory of their own, no coherent system and were philosophically inconsistent, thereby encouraging the view of a theory-free experimental practice that sounds at odds with the project of describing exclusively their philosophical notions.

Siegfried's assumption that "chemical composition has had no fundamental conceptual change" since Lavoisier and Dalton suggests that there is no history of chemical atomism since Dalton's atomic theory. Furthermore, the book's title, *From Elements to Atoms*, suggests that the idea of the element is now obsolete and that modern chemistry focuses exclusively on the atom. This is a view that many chemistry teachers will no doubt debate. ■

Bernadette Bensaude-Vincent is in the Department of Philosophy, Université de Paris X, 92001 Nanterre Cedex, France.



Antoine-Laurent Lavoisier revolutionized chemistry in the eighteenth century.

Tuberculosis hits back

Return of the White Plague: Global Poverty and the 'New' Tuberculosis

edited by Matthew Gandy & Alimuddin Zumla

Verso: 2003. 320 pp. \$35, £25, Can\$51

D. A. Mitchison

People tell me that tuberculosis has come back, to which I reply: "It never went away." In the 1970s it was thought that tuberculosis was no longer a threat, at least in Britain, and would soon be eradicated. That this view was mistaken is illustrated in this book by a series of essays describing major failures in control and the hotspots where these failures are leading to the emergence of drug-resistant strains.

The World Health Organization estimates that there is steady but slow progress towards better control of tuberculosis worldwide. This progress could be accelerated by the widespread adoption of the DOTS strategy, a five-prong policy comprising government commitment, diagnosis by sputum smears, short-course chemotherapy with supervision of drug-taking, regular drug supplies, and sufficient record-keeping to allow progress to be analysed. An increasing number of countries across the world are adopting this strategy.

In contrast to this guarded optimism, the descriptions of the hotspots in *Return of the White Plague* paint a frightening picture. Dormant tuberculosis infections are frequently reactivated by HIV infection with devastating results. AIDS kills a sizeable proportion of the population in sub-Saharan Africa, often with terminal tuberculosis. The spread of HIV in these countries has been accelerated by political blindness at the highest level. The exception is Uganda, which was quick to recognize the need to explain the origin of HIV infection and how preventive methods could be used and, as a result, has successfully controlled its spread. Recent agreements by pharmaceutical companies to provide generic anti-retroviral drugs have raised hopes, although it is uncertain whether these can be used effectively in the huge populations afflicted with HIV. An effective vaccine is still far in the future.

The chapter on the outbreak of drug-resistant tuberculosis in New York is marred by the gloomy prognostication that better treatment would not be an effective control measure. In fact, the control policy of improving treatment with compulsory segregation of patients with resistant strains, carried out by Paula Fujiwara, who ran the Bureau of Tuberculosis Control in New York, has been highly effective, although at its



Locked up with a killer: drug-resistant tuberculosis is a widespread problem in Russia's prisons.

peak it cost \$50 million per year. The high incidence of tuberculosis in the poorest and most socially excluded populations in London cries out for a more determined and actively led control programme like that carried out in New York.

There are also chapters on drug-resistant tuberculosis in Haiti, Peru and, most frighteningly, in the prisons of the former Soviet bloc. The main reason for the Soviet problem was the failure of the authorities there to make any contact with those developing the successful treatment of the disease between 1950 and 1980. No clinical trial of the treatment was ever carried out in the Soviet Union. This political isolation led to the use of inappropriate regimens of treatment with the result that drug resistance became widespread. These core chapters of the book make us reflect on the main causes of the resurgence of tuberculosis: HIV infection; migration, which spreads HIV as well as tuberculosis; a lack of proper scientific leadership in control policies; and finally, and perhaps most importantly, poverty.

Other chapters are of mixed value and concentrate too much on failure rather than on possible progress. There are also notable omissions. For example, the introductory description of the tuberculous lesions only describes part of the process. The haematogenous spread from the primary site of infection to the adult lesions is not described at all, yet this is key to understanding why HIV infection results in such a high

rate of reactivation of tuberculosis and how the BCG vaccine works in preventing the spread. Similarly, the chapter on immigration obscures the evidence that immigrants initially carry with them the tuberculosis incidence rate in the country from which they came.

Finally, chapters on advocacy and action include a good, albeit long-winded, account of the importance of doctors understanding their patients. The chapter on the role of science in tuberculosis control makes no mention of issues such as the duration of antigen stimulation, which is key to the understanding of DNA vaccines, nor to the development of short-course chemotherapy, or of the need for new drugs to shorten treatment and the current emphasis on using genomics in drug design.

The chapter ends with the hope engendered by DNA vaccines. Yet the UK Medical Research Council (MRC) has failed to support work in this area, despite recommendations from its own review committee that a clinical unit be established. Since 1986, support for basic research into tuberculosis in Britain has declined steadily, and there is now only a fragmented unit working at the MRC's lab at Mill Hill near London. How can the MRC defend its failure to support research on an infection of such worldwide importance?

D. A. Mitchison is in the Department of Medical Microbiology, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK.

Gardening by numbers

Architect Charles Jencks' Garden of Cosmic Speculation.

John D. Barrow

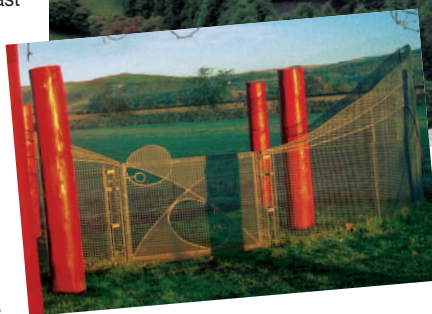
Physics is much like gardening. Both create small worlds that model some of the most important and appealing parts of reality — physicists on paper using mathematical patterns and judicious approximation, gardeners by reconnecting the real. Both look for order amid complexity. Both make the global local; one uses reductionism, the other, like all art, does not. The two rarely meet, but when they do, it is either to help us escape from a maze or to lead us into the temptations of fantasy.

Jorge Luis Borges' Garden of the Forking Paths, for instance, tells us about a multiverse of bifurcating realities where all possibilities are realized. Another example has been created by the architect Charles Jencks. Over the past eight years, Jencks has created a 12-hectare Garden of Cosmic Speculation at Portrack in Scotland, UK. It is not yet open to the public, other than by special arrangement, so most people will have to limit their experience of this monumental creation to Jencks' sumptuously illustrated book *The Garden of Cosmic Speculation* (Frances Lincoln, 2003), in which he describes how the garden was conceived, designed and made, in memory of his late wife, Maggie Keswick, who played a key role in its early design and landscaping.

In the kitchens of the house, the panelling designs illustrate the period doubling of the logistic mapping as it tumbles towards chaos. There are huge carpets with beautiful designs of non-periodic Penrose tiling in the guest rooms; spirals and topological surfaces indoors and out; and asymmetrical room designs that display the tidal effects of passing gravitational waves, wave forms, catastrophe surfaces, platonic solids and broken symmetries. These create a spectacular interior design that respects the Scottish style, using muted cream colours and appealing shapes that all have a deeper mathematical message — a sort of quantum Rennie Mackintosh — artfully rendered by carpenter Bobby Dixon.

Outside, things are bigger and bolder. Crescents and strips of water look at first like mere ornamental lakes but on closer inspection they can be seen to trace the characteristic patterns of the phase space of trajectories in the famous Henon strange attractor. This is like the work of artist Andy Goldsworthy writ large, except that it is permanent. All that's missing is for the watching rabbits to multiply in Fibonacci sequence.

My personal favourite items are the beautiful wrought-iron gates, made by John Gibson, that are found all over the estate. They are highly original in concept and quite stunning in the simplicity of their execution. Some show simple wave interactions, with solitary waves passing through each other in metal like starships in the night.



Square roots? The gardens at Portrack House in Scotland are based on mathematical designs.

Occasionally, the iron fetters yield to lettering, again with a (Möbius) twist; ambigrammi abound, some reading the same backwards, some upside-down. This elegant calligraphy characterizes the complementary work in stone.

Dry-stone walls, themselves fractals, have been built by Hugh Drysdale with sinusoidally undulating tops to enclose the inner garden. On the ground there are networks of stone patterns that capture the weave of Scottish tartans, metamorphosing between order and disorder as their phase changes. The doors to the garden each define an open fretwork pattern that creates the illusion of looking down an endless hall of mirrors.

Beyond the inner garden are works of sculpture in wood and stone that celebrate the great equations of physics and the parts of the Universe that they govern. These modern tablets of the natural law run as a frieze along the tops of the glass-houses — the timeless formulae of Einstein, Lorentz, Lagrange and Schrödinger stretch out like the score of a great symphony. Below them sit three great celestial globes of bronze depicting the babylonian, ptolemaic and platonic world views. The sculptures are completed by three more brazen displays of our conceptions of the Universe, life and the atomic world. On their bounding circles are etched the equations that govern the quantum structure of the atom and the force of gravity. Here I detected one scientific 'error' in the artistic work: Einstein's equations of general relativity contain the wrong power of the speed of light.

The outer reaches of the garden display topographical manifestations of curved space and black-hole horizons, vast wire sculptures sprouting into the air above the water to recreate the cascades of bubble-chamber interactions of subatomic particles, and there is a fine collection of non-euclidean garden furniture. Life emerges from the ground in the form of helices in wood, stone and metal, doubling back to form steps and spiral staircases that evoke the genetic code as they twist atoms into life.

Finally, there is a path across a broad, green-banked lake. It leads to a white, stone staircase that rises steeply, narrowly at first, then broadens suddenly to zig-zag upwards towards a white, angular building high above. The path upwards is a timeline for the history of the Universe: expanding, then suddenly inflating, producing stars, planets and life over billions of years. Along the path are strange cosmic milestones, some natural, some fashioned, that reflect the antiquity and activity of the Universe's long and eventful past — catastrophic, volcanic, episodic — that brought us from simplicity to complexity, from vacuum to life.

Jencks' garden of delights inspires a feeling for the complexity and beauty of the unseen Universe, and for the simple laws and symmetries behind the complexities and asymmetries we see around us. As we enter it, we feel a fresh affinity with a nature that is larger than life, taking a step into an unspoiled Garden of Eden.

John D. Barrow is in the Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge CB3 0WA, UK. His latest book, *The Constants of Nature*, is published by Vintage.

A meeting with Enrico Fermi

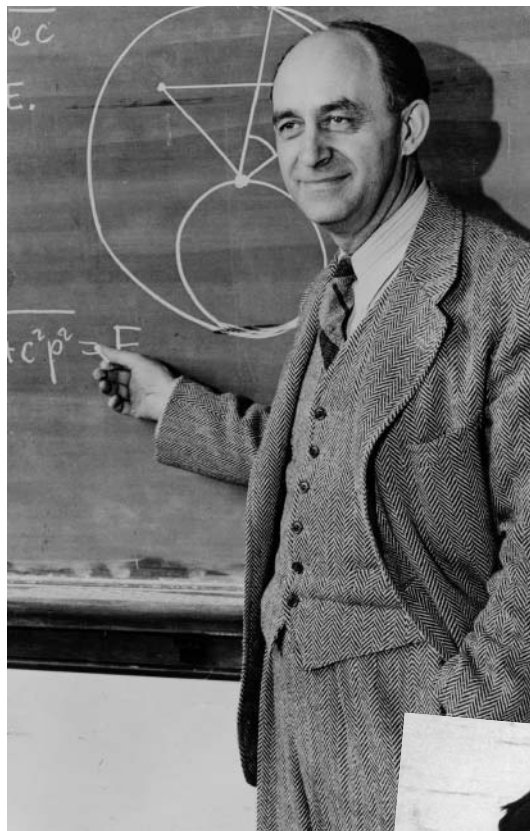
How one intuitive physicist rescued a team from fruitless research.

Freeman Dyson

One of the big turning points in my life was a meeting with Enrico Fermi in the spring of 1953. In a few minutes, Fermi politely but ruthlessly demolished a programme of research that my students and I had been pursuing for several years. He probably saved us from several more years of fruitless wandering along a road that was leading nowhere. I am eternally grateful to him for destroying our illusions and telling us the bitter truth.

Fermi was one of the great physicists of our time, outstanding both as a theorist and as an experimenter. He led the team that built the first nuclear reactor in Chicago in 1942. By 1953 he was head of the team that built the Chicago cyclotron, and was using it to explore the strong forces that hold nuclei together. He made the first accurate measurements of the scattering of mesons by protons, an experiment that gave the most direct evidence then available of the nature of the strong forces.

At that time I was a young professor of theoretical physics at Cornell University, responsible for directing the research of a small army of graduate students and postdocs. I had put them to work calculating meson–proton scattering, so that their theoretical calculations could be compared with Fermi's measurements. In 1948 and 1949 we had made similar calculations of atomic processes, using the theory of quantum electrodynamics, and found spectacular agreement between experiment and theory. Quantum electrodynamics is the theory of electrons and photons interacting through electromagnetic forces. Because the electromagnetic forces are weak, we could calculate the atomic processes precisely. By 1951, we had triumphantly finished the atomic calculations and were looking for fresh fields to conquer. We decided to use the same techniques of calculation to explore the strong nuclear forces. We began by calculating meson–proton scattering, using a theory of the strong forces known as pseudoscalar meson theory. By the spring of 1953, after heroic efforts, we had plotted theoretical graphs of meson–proton scattering. We joyfully observed that our calculated numbers agreed pretty well with Fermi's measured numbers. So I made an appointment to meet with Fermi and show him our results. Proudly, I rode the Greyhound bus from Ithaca to Chicago with



Crossed paths: A discussion with Enrico Fermi (above) made Freeman Dyson (right) change his career direction.

a package of our theoretical graphs to show to Fermi.

When I arrived in Fermi's office, I handed the graphs to Fermi, but he hardly glanced at them. He invited me to sit down, and asked me in a friendly way about the health of my wife and our newborn baby son, now fifty years old. Then he delivered his verdict in a quiet, even voice. "There are two ways of doing calculations in theoretical physics", he said. "One way, and this is the way I prefer, is to have a clear physical picture of the process that you are calculating. The other way is to have a precise and self-consistent mathematical formalism. You have neither." I was slightly stunned, but ventured to ask him why he did not consider the pseudoscalar meson theory to be a self-consistent mathematical formalism. He replied, "Quantum electrodynamics is a good theory because the forces are weak, and when the formalism is ambiguous we have a clear physical picture to guide us. With the pseudoscalar meson theory there is no

physical picture, and the forces are so strong that nothing converges. To reach your calculated results, you had to introduce arbitrary cut-off procedures that are not based either on solid physics or on solid mathematics."

In desperation I asked Fermi whether he was not impressed by the agreement between our calculated numbers and his measured numbers. He replied, "How many arbitrary parameters did you use for your calculations?" I thought for a moment about our cut-off procedures and said, "Four." He said, "I remember my friend Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk." With that, the conversation was over. I thanked Fermi for his time and trouble, and sadly took the next bus back to Ithaca to tell the bad news to the students. Because it was important for the students to have their names on a published paper, we did not abandon our calculations immediately. We

finished them and wrote a long paper that was duly published in the *Physical Review* with all our names on it. Then we dispersed to find other lines of work. I escaped to Berkeley, California, to start a new career in condensed-matter physics.

Looking back after fifty years, we can clearly see that Fermi was right. The crucial discovery that made sense of the strong forces was the quark.

Mesons and protons are

little bags of quarks. Before Murray Gell-Mann discovered quarks, no theory of the strong forces could possibly have been adequate. Fermi knew nothing about quarks, and died before they were discovered. But somehow he knew that something essential was missing in the meson theories of the 1950s. His physical intuition told him that the pseudoscalar meson theory could not be right. And so it was Fermi's intuition, and not any discrepancy between theory and experiment, that saved me and my students from getting stuck in a blind alley. ■

Freeman Dyson is at the Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA.



A problem of distance

Bohdan Paczynski

How far is the Pleiades star cluster from Earth? The latest measurement suggests that there is a problem with data from the Hipparcos satellite, which will have repercussions for estimating other astronomical distances.

The European Space Agency's satellite Hipparcos¹ has enjoyed a successful astrometric mission, measuring the distances and motions of more than 100,000 nearby stars. But its measurement of the distance to the well-known star cluster the Pleiades (or Seven Sisters; Fig. 1) has caused a major controversy². The cluster is more than 100 parsecs from Earth, equivalent to a distance of 300 light years. But exactly how much more than 100 parsecs is a matter of dispute. Xiaopei Pan and colleagues³ (page 326 of this issue) offer a plausible resolution.

The oldest direct method of measuring the distance to a star is through 'parallax', the small shift in a star's position in the sky caused by the Earth's orbital motion. From this comes the name of a unit of distance commonly used by astronomers — the parsec (pc), which corresponds to roughly 3×10^{13} kilometres. At a distance of 100 pc, the parallax is only 0.01 arcseconds (another common unit: an arcsecond is 1/3,600 of a degree). But the Earth's atmosphere blurs stellar images, such that, from the planetary surface, it is difficult to measure parallax with an accuracy better than 0.01 arcseconds. So although the parallax technique has been used⁴, not everybody accepts the measurements as an accurate distance for the Pleiades.

Traditionally, the distance to the Pleiades cluster has instead been estimated using a method called main-sequence fitting. This takes advantage of the empirical relation between the colour of a star and its brightness, which is theoretically well understood for stars in the main sequence (stars powered by the fusion of hydrogen into helium are 'main sequence', and remain so for roughly 90% of their observable lives). First, the distance to the Hyades cluster, the nearest star cluster to Earth, was measured directly through simple geometrical techniques. Then the main-sequence stars in Pleiades were compared with those in Hyades. Because the apparent brightness of a star is inversely proportional to the square of its distance, the distance to Pleiades was worked out⁵ at 132 ± 4 pc.

Enter Hipparcos. In orbit above the Earth's atmosphere, this satellite has provided measurements of parallax that are accurate to 0.001 arcseconds, a major breakthrough. But for stars at a distance of 132 pc, even Hipparcos could not measure parallaxes to an accuracy better than 13%. However, by



Figure 1 How far? The Pleiades star cluster, also known as the Seven Sisters, is more than 300 light years from Earth. Exactly how far away the cluster is, however, is controversial. A measurement by the Hipparcos satellite conflicts with an estimate based on stellar brightness. A new calculation by Pan *et al.*³ suggests that the Hipparcos data may be flawed.

combining the measurements made of many stars in the Pleiades cluster, that accuracy could be improved, and its distance was announced⁶ to be 118 ± 4 pc — a serious conflict with the distance obtained from main-sequence fitting.

The consequences of this discrepancy are profound. If Hipparcos is right and the main-sequence-fitting result is wrong, then not only are the distances that have been calculated to stellar globular clusters wrong, but we are also wrong about the estimates of cluster ages and a seemingly robust way to measure the size of the Universe. On the other hand, the Hipparcos measurements may have suffered some unknown systematic error, which should be uncovered lest a similar error affect the future, more accurate astrometric mission called Gaia⁷.

Many attempts have been made to resolve the dispute, but none has succeeded. This latest attempt, by Pan *et al.*³ using the Palomar Testbed Interferometer in California, is a major step towards a firm distance determination, although I doubt that it will convince hard-core sceptics. The principle of

the method is simple and elegant. One of the brightest stars in the Pleiades cluster, named Atlas, is a binary system. The two stars move around each other under gravity, with an orbital period of 290.8 days. Pan *et al.* measured the semi-major axis of this orbit to be 0.0129 ± 0.0001 arcseconds, a measurement that is more robust than parallax.

Ideally, the next step (which is missing so far) would be to measure the radial velocities of the two stars. If the changes in radial velocity are combined with the orbital period and the angular size of the orbit, a direct measure of the distance to the system is obtained. Unfortunately, the two stars rotate so rapidly that it is difficult to measure their radial velocities (their absorption lines overlap). But it is not impossible. In fact, another group is working on such measurements, and once that work is complete I expect the controversy to be resolved — to the satisfaction of some but the discomfort of others.

Pan *et al.*³ do not have those radial velocities, so they have used a less direct, but still simple, method. In a nutshell, it is based on two relations — Kepler's third law and the

mass–luminosity relation — that connect the unknown distance of the Pleiades cluster with the mass of the Atlas binary. The authors conclude that the distance is 135 ± 2 pc. So if Pan *et al.* are right, Hipparcos is wrong.

I suspect that Hipparcos enthusiasts will not accept this measurement: using the mass–luminosity relation is equivalent to using main-sequence fitting, which they claim has problems. But I think Pan *et al.* have got it right. There could well be a systematic error in the Hipparcos data, owing to its highly eccentric orbit. This orbit was not planned but was caused by a failure of some of the satellite's boosters shortly after launch. Initially it seemed that the mission was lost, but thanks to the ingenuity of the Hipparcos team, the satellite was still able to complete its task. I would not be surprised, though, if there turns out to be some residual problem with the defective orbit that has caused an unfortunate shift in the Hipparcos data⁸. If

we are to have confidence in our measurements of astronomical distances, not just the distance to the Pleiades cluster, this problem must be solved. ■

Bohdan Paczynski is in the Department of Astrophysics, Princeton University, Princeton, New Jersey 08544-1001, USA.
e-mail: bp@astro.princeton.edu

1. Perryman, M. A. C. *et al.* *Astron. Astrophys.* **323**, L49–L52 (1997).
2. Mermilliod, J.-C., Turon, C., Robichon, N., Arenou, F. & Lebreton, Y. in *ESA SP-402: Hipparcos Venice '97* (eds Perryman, M. A. C. & Bernacca, P. L.) 643–650 (European Space Agency, Noordwijk, 1997).
3. Pan, X., Shao, M. & Kulkarni, S. R. *Nature* **427**, 326–328 (2004).
4. Gatewood, G., Kiewiet de Jonge, J. & Han, I. *Astrophys. J.* **533**, 938–943 (2000).
5. Meynet, G., Mermilliod, J.-C. & Maeder, A. *Astron. Astrophys. Suppl. Ser.* **98**, 477–504 (1993).
6. van Leeuwen, F. *Astron. Astrophys.* **341**, L71–L74 (1999).
7. Perryman, M. A. C. in *ESA SP-485: Proc. 1st Eddington Workshop on Stellar Structure and Habitable Planet Finding* (eds Favata, F., Roxburgh, I. W. & Galadi, D.) 229–234 (European Space Agency, Noordwijk, 2002).
8. Narayanan, V. K. & Gould, A. *Astrophys. J.* **523**, 328–339 (1999).

Cell division

Burning the spindle at both ends

Rebecca W. Heald

Accurate transmission of the genome during cell division requires the physical separation of replicated chromosomes. The identities of two molecular motors needed to do the job in fruitflies are now revealed.

A dramatic event in the life of a cell is its transformation into two genetically identical progeny. This is achieved during mitosis, when an exact complement of chromosomes is partitioned to each half of the cell, just before it pinches into two¹. Errors in this process can result in cell death or contribute to cancer. The mitotic spindle — the apparatus that distributes the chromosomes — has been studied for decades. Nonetheless, the molecular mechanisms underlying chromosome transport have remained elusive. On page 364 of this issue², however, Rogers *et al.* describe two related motor proteins that are essential for spindle dynamics and chromosome segregation in fruitflies.

The events of cell division require dynamic elements of the cell's internal skeleton, which assemble into macromolecular structures capable of performing work. The spindle is one such structure: this highly dynamic yet ordered assemblage is composed of cytoskeletal elements known as microtubule filaments — along with many associated proteins — which form a bipolar array³ (Fig. 1). Microtubules are polymers made up of tubulin proteins, and they grow or shrink when tubulin is added or lost from their ends. These ends are termed 'plus' and 'minus' to distinguish their behaviours. Minus ends are focused at each pole of the spindle, emanating from a nucleating struc-

ture called the centrosome. Plus ends grow outwards from the poles, capturing and moving chromosomes or overlapping at the centre of the spindle. Several classes of microtubule-based motor proteins are also required for chromosome segregation. Each type of motor moves in a set direction along microtubules. Together they cross-link and sort the microtubules according to their structural polarity, and mediate chromosome interactions with the spindle¹.

Before embarking on mitosis, a cell duplicates its chromosomes, producing pairs of 'sister chromatids'. The members of each pair are identical. In what is known as prometaphase of mitosis, these chromatids become tethered to the spindle, such that one member of each pair is attached to a bundle of microtubules emanating from one pole, and the other member is connected to the other pole. 'Kinetochore' are the microtubule landing pads on chromosomes³. A tug-of-war ensues, driven by the addition or loss of tubulin (that is, microtubule polymerization or depolymerization) at the kinetochores. This tug-of-war results in chromosomes becoming aligned during metaphase of mitosis, such that each sister faces its final destination (Fig. 1a). Finally, anaphase begins when the glue holding sisters together is dissolved, and they take off towards opposite spindle poles (Fig. 1b).

Long-standing questions surround the

molecular mechanism of kinetochore-mediated chromosome alignment and segregation. By marking a segment of the spindle microtubule lattice and watching what happens to it as sister chromatids move apart, researchers have described two phenomena in which microtubule depolymerization is coupled to chromosome separation in anaphase (Fig. 1c, d). First, the marked region moves polewards, indicating that depolymerization is occurring at the minus ends of the microtubules; this could serve to reel in the chromosomes attached at the plus ends^{4,5}. (Interestingly, this poleward 'flux' occurs during metaphase as well, but at this time the minus-end depolymerization is balanced by plus-end polymerization.) The second observed phenomenon is the disappearance of the labelled region as the chromosome moves past it, indicating that the kinetochores are chewing up microtubule plus ends as they move to the pole — a mechanism known as Pac-Man in reference to the video game^{6,7}.

But what are the molecules that drive these depolymerization events during anaphase? Rogers and colleagues² provide some answers, through a functional examination of two motor proteins in cultured fruitfly cells and early fruitfly embryos. The kinesin-like proteins KLP59C and KLP10A belong to a specialized class of microtubule-based motors (the Kin I class, for 'kinesin internal catalytic domain') that do not move along microtubules but instead bind to their ends and induce depolymerization⁸. During anaphase, KLP59C is found at kinetochores, whereas KLP10A is predominantly at spindle poles. Rogers *et al.* find that inhibiting either of these proteins interferes with chromosome segregation. By monitoring the effects on spindle microtubules, they show that KLP10A is required for microtubule depolymerization at spindle poles — that is, for the 'reeling in' mechanism. Meanwhile, KLP59C is the molecular bite behind Pac-Man (Fig. 1c, d). Interfering with these motors also causes defects in chromosome positioning during metaphase, suggesting that similar mechanisms contribute to chromosome movement during alignment.

The Kin I kinesins were logical candidates for driving chromosome dynamics through microtubule depolymerization, and have been studied previously in other experimental systems. So it is somewhat surprising that a definitive role for these proteins in anaphase has not been demonstrated before now. One explanation might be that fruitfly embryos have not been used before to study the contribution of these proteins — yet fruitfly embryos provide a unique and powerful system with which to view spindle dynamics. Their relatively large size permits analysis by high-resolution microscopy, and their rapid cell cycle does not include the quality-control 'checkpoints' found at later stages of fruitfly

development and in other organisms, whereby mitotic progression is delayed if there are defects in chromosome alignment during metaphase. So, when Rogers *et al.* deactivated KLP10A and KLP59C, the chromosomes lined up abnormally — but compensatory mechanisms were not engaged, and effects on anaphase could still be assessed.

Another possible explanation is that members of the Kin I family also control other aspects of microtubule dynamics. So inhibiting them can give rise to marked structural defects in the spindle, producing indirect effects on chromosome movements and making it difficult to identify a specific anaphase role for the motors. And, interestingly, although functional characterization of a vertebrate Kin I relative revealed anaphase defects⁹, detailed analysis indicated that its primary role at the kinetochore is to depolymerize incorrectly attached microtubules, thus preventing the aberrant connections that cause lagging chromosomes and mis-segregation¹⁰. As multicellular organisms that are more complex than fruitflies possess multiple Kin-I-related proteins, it could be that the true functional counterparts of the anaphase fruitfly motors have yet to be characterized.

What remains to be discovered? One question is how microtubules maintain their attachment to kinetochores and spindle poles while undergoing polymerization and depolymerization. Also, what do motile motors at the kinetochore — such as dynein and CENP-E — contribute to the process? How are the multiple microtubules that attach to a single chromatid coordinately regulated? Mitosis provides a rich source of questions for the mechanistically inquisitive.

Rebecca W. Heald is in the Department of Molecular Cell Biology, University of California at Berkeley, Berkeley, California 94720-3200, USA.
e-mail: heald@socrates.berkeley.edu

1. Scholey, J. M. *et al.* *Nature* **422**, 746–752 (2003).

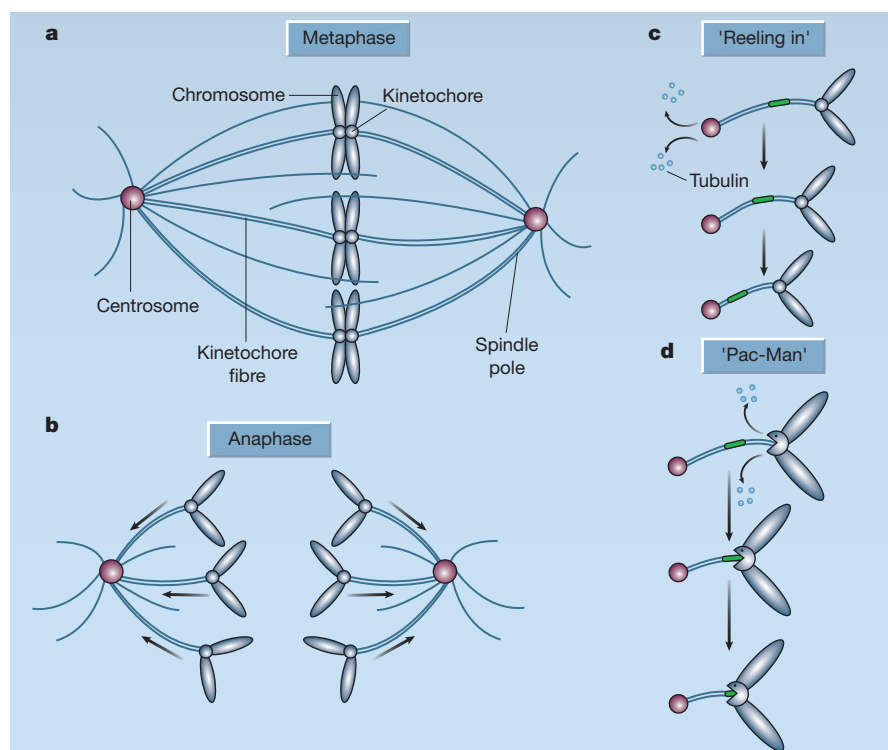


Figure 1 Achieving chromosome segregation in fruitfly cells. **a**, In metaphase, sister chromosomes are already attached via their kinetochores to the plus ends of kinetochore fibres (bundles of microtubule filaments), in the centre of the mitotic spindle. Sisters become aligned and oriented towards opposite spindle poles, where microtubule minus ends are focused at centrosomes. **b**, In anaphase, chromosomes move apart along kinetochore fibres, as the microtubules are depolymerized. **c**, **d**, Two mechanisms of chromosome movement have been proposed, based on experiments in which a marked segment (green) of the kinetochore fibre is tracked. **c**, In the 'reeling-in' mechanism^{4,5}, the minus end of the kinetochore fibre is depolymerized, while the chromosome maintains attachment at the plus end. **d**, In the 'Pac-Man' mechanism^{6,7}, the kinetochore fibre is chewed up from the plus end; however, the chromosome remains attached and so moves polewards. Rogers *et al.*² have discovered that the motor protein KLP10A is behind the reeling-in mechanism, and KLP59C is the Pac-Man.

2. Rogers, G. *et al.* *Nature* **427**, 364–370 (2004).

3. Rieder, C. L. & Salmon, E. D. *Trends Cell Biol.* **8**, 310–318 (1998).

4. Mitchison, T. J. *J. Cell Biol.* **109**, 637–652 (1989).

5. Brust-Mascher, I. & Scholey, J. M. *Mol. Cell. Biol.* **13**, 3967–3975 (2002).

6. Mitchison, T. J., Evans, L., Schulze, E. & Kirschner, M. *Cell* **45**, 515–527 (1986).

7. Gorbisky, G. J., Sammak, P. J. & Borisy, G. G. *J. Cell Biol.* **104**, 9–18 (1987).

8. Desai, A., Verma, S., Mitchison, T. J. & Walczak, C. E. *Cell* **96**, 69–78 (1999).

9. Maney, T. *et al.* *J. Cell Biol.* **142**, 787–801 (1998).

10. Kline-Smith, S. L., Khodjakov, A., Hergert, P. & Walczak, C. E. *Mol. Biol. Cell* doi:10.1091/mbc.E03-08-0581 (2003).

Nanotechnology

Dreams of a hollow future

Luis Hueso and Neil Mathur

Carbon nanotubes have become familiar components in nanotechnology. Nanotubes made from inorganic materials are now on the rise, the latest creation being nanoscale tubes of a complex manganese oxide.

Fabricating small structures has long been fashionable in physics. The rationale is that reducing one or more dimensions of a system below some key length scale can change the system's behaviour — carbon nanotubes are a good example. But nanotubes made from other materials are also proving useful for technological applications. In *Applied Physics Letters*, Levy *et al.*¹ add to the catalogue with

their report of the growth of nanotubes made of a manganese oxide, namely a manganite.

Carbon nanotubes, discovered by Iijima² in 1991, can be thought of as rolled-up sheets of carbon atoms. The tubes have diameters as small as one nanometre, and are typically several micrometres long. Thus they are, effectively, one-dimensional. This reduced dimensionality creates a new playground for

physicists, where the conventional description of the electronic structure of three-dimensional materials breaks down³. But interesting effects are not restricted to only the smallest nanotubes. Crude carbon-nanotube structures, consisting of imperfectly concentric cylinders with diameters as large as a few hundred nanometres, also have technological uses. The high aspect ratio of these structures means that electrons can be emitted easily from their tips. If these electrons then traverse a vacuum and excite a phosphor on a screen, this forms the basis of a display pixel. Indeed, proof-of-principle displays using such multi-wall nanotube structures have been fabricated and promise to be ten times more energy efficient than competing plasma technology⁴.

The techniques of modern materials

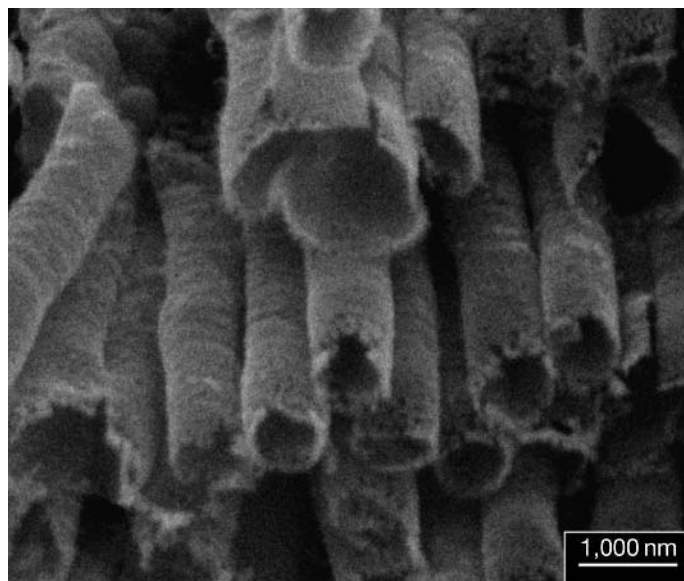


Figure 1 Going inorganic. Levy and colleagues¹ made these inorganic nanotubes from lanthanum–praseodymium–calcium manganite. The changing electrical resistance of this material in a magnetic field suggests that such nanotubes could be usefully applied in nanotechnological devices, such as fuel cells.

science also allow the fabrication of inorganic tubular nanostructures. As with the multi-wall carbon nanotubes, no key length scales are probed, but there is, again, the promise of technological applications. Good examples are the piezoelectric nanotubes made from complex oxides such as barium titanate^{5,6} and strontium–bismuth tantalate⁷. ‘Piezoelectric’ means that these polycrystalline tubes can be strained when an electrical voltage is applied, and vice versa. Each tube could be triggered individually to release a small quantity of ink for ink-jet printing, or to deliver drugs into a patient. Sensor, actuator and data-storage applications are also possible.

The excitement generated by piezoelectric nanotubes has now inspired Levy *et al.*¹ to emulate the same growth technique using a different and resurgent class of oxides. Manganites are complex oxides that adopt a pseudo-cubic perovskite crystal structure. Half a century ago, it was found that an applied magnetic field could significantly change the electrical resistance of these materials⁸, but it is only in the past decade that these ‘magnetoresistance’ effects have been studied in detail. The catalyst for this activity was the discovery of colossal magnetoresistance in a thin film⁹, just as thin-film magnetoresistance effects were making the transition from the laboratory to application in read heads for computer disk drives.

To fabricate their nanotubes of lanthanum–praseodymium–calcium manganite, Levy *et al.* first made a porous template by chemically etching films of mylar and polycarbonate that had been bombarded with heavy ions. They then introduced a precursor solution into the (wetted) pores, and achieved crystallization by heating the template. Microstructures comprising long, thin-walled nanotubes formed spontaneously (Fig. 1). Through various structural characterization techniques, Levy *et al.* confirmed

that each tube is composed of manganite nanocrystals. Moreover, rough estimates of the magnetic properties match those expected for bulk samples of this manganite.

How might manganite nanotubes impact on technology? One possible application is in solid-oxide fuel cells. A fuel cell differs from a battery in that reactants may be continuously fed into it and exhausted. The microstructure demonstrated by Levy *et al.* immediately suggests a means by which gases may be efficiently distributed in such a cell. And as manganites conduct both electrons and oxygen ions, and are resistant to high-temperature oxidizing environments, they make good cathodes.

More speculatively, nanotubes made from metallic manganites could act as highly localized sources of electrons possessing spins of a particular orientation. This is possible because the spins of the conduction electrons in manganites can be aligned perfectly, whereas in ordinary magnetic metals such as cobalt the alignment is only partial. It is possible to imagine the nanoscale engineering of electronic circuits in which the spin of electrons, as well as their charge, could be manipulated with precision — a valuable capability for spin-sensitive scanning probe microscopy, and perhaps, ultimately, quantum computing.

Nanotube structures may also offer a means of tuning the strong interactions that exist between the magnetic, electronic and crystal structures of a manganite. These interactions generate rich phase-coexistence phenomena over a wide range of length scales, as has been revealed by imaging methods¹⁰. For example, a ferromagnetic metallic phase may coexist with an antiferromagnetic insulating phase. In a nanotube, the delicate balance between the diverse phases could be tuned readily through the stresses associated with the unconventional geometry. Exploring the parameter space of chemical



100 YEARS AGO

Elementary Physiology and Hygiene. This book has obviously been written to supply the wants of the American schoolchild, and consequently “the subject of alcohol has been treated very thoroughly and in full compliance with the laws of the various States.” “Throughout the book the effects of alcohol and other narcotics have been discussed in close connection with the accounts of the functions of the body.” The above quotations from the author’s preface show that it has been a pleasure to him to comply in his book with the law enjoining that all text-books of physiology used in American State schools must contain a description of the effects of alcohol on the body. So thoroughly has this instruction been carried out that it appears on reading the book as if in many cases the very brief descriptions of the physiology of the different tissues had been written chiefly as introductions in order to make clear the dire effects of alcohol, which are subsequently described in each case... Truly this book must be appalling reading to the American schoolchild whose parents may be in the habit of making even moderate use of alcoholic drinks.

From *Nature* 21 January 1904.

50 YEARS AGO

For the Seventh Arthur Stanley Eddington Memorial Lecture... Prof H. H. Price, Wykeham professor of logic in the University of Oxford, spoke on “Some Aspects of the Conflict between Science and Religion”... Prof. Price argues that of late the main conflict between science and religion has been over two opposed conceptions of human personality, a materialist and a dualist one. On one side, it is held that all mental processes are produced by and inseparable from bodily processes, or else actually are bodily. According to the religious conception, on the other side, some kind of cognition of the divine is possible, which cannot be supposed to have any ordinary physiological correlates, and also some kind of other-worldly existence distinct from the present bodily one. The systematic investigation of the phenomena of para-normal cognition, such as telepathy, clairvoyance and precognition, has made the materialist conception far less plausible, if not untenable. A dualist type of theory... can no longer be dismissed as unscientific and superstitious.

From *Nature* 23 January 1954.

composition, grain size, tube dimensions and tube distribution should reveal more exciting possibilities ahead. The future of nanotubes looks anything but hollow. ■

Luis Hueso and Neil Mathur are in the Department of Materials Science and Metallurgy, University of Cambridge, New Museums Site, Pembroke Street, Cambridge CB2 3QZ, UK.

e-mails: leh29@cam.ac.uk
ndm12@cus.cam.ac.uk

1. Levy, P., Leyva, A. G., Troiani, H. E. & Sánchez, R. D. *Appl. Phys. Lett.* **83**, 5247–5249 (2003).
2. Ijima, S. *Nature* **354**, 56–58 (1991).
3. Ishii, H. et al. *Nature* **426**, 540–544 (2003).
4. Amarantunga, G. *IEEE Spectrum* **40**, 28–32 (2003).
5. Hernandez, B. A. et al. *Chem. Mater.* **14**, 480–482 (2002).
6. Luo, Y. et al. *Appl. Phys. Lett.* **83**, 440–442 (2003).
7. Morrison, F. D., Ramsay, L. & Scott, J. F. J. *Phys. Condens. Matter* **15**, L527–L532 (2003).
8. Volger, J. *Physica* **20**, 49–54 (1954).
9. Jin, S. et al. *Science* **264**, 413–415 (1994).
10. Mathur, N. & Littlewood, P. *Physics Today* **56**, 25–30 (2003).

Psychology

Insight and the sleep committee

Pierre Maquet and Perrine Ruby

We all spend about a third of our lives asleep, an essential but seemingly unproductive state. Experimental evidence now emerges to support anecdotal evidence that sleep can stimulate creative thinking.

Does this experience seem familiar? The solution to an unfathomable problem, left unresolved in the evening, effortlessly pops into your mind the following morning. Although many people believe that sleep plays a role in these flashes of insight, this is a hypothesis that has not been rigorously tested. On page 352 of this issue¹, however, Wagner and collaborators provide evidence that sleep can have a beneficial effect on insight.

The authors have applied a clever test that allows them to determine exactly when insight occurs in the time course of learning².

In this task, the participants have to transform a string of eight digits into a new string, the last digit of which is the final solution (see Fig. 1 on page 353). To do this, they are instructed to apply two simple rules sequentially, from one digit to the next. However, unknown to the subjects, another rule is hidden in the material: the last three responses mirror the three preceding ones. Discovering the hidden rule can speed up the execution of the task, as the final solution is known when the third digit is specified.

All participants were trained in the task (three blocks of tasks), then retested (for ten

blocks) after eight hours. During the eight-hour period the subjects were either awake (during daytime or during the night) or sleeping. When they were retested, the proportion who gained insight in those allowed to sleep (60%) was more than twice that in those who remained awake (22%). If subjects were exposed to the task continuously for 13 blocks, without having been trained the day before, the proportion who gained insight was similar to that in the awake groups. In other words, the favourable effect of sleep on insight occurred only if a memory had been formed before the sleep period.

The data further suggest that the conscious use of the hidden rule did not evolve from procedural learning — that is, from the unconscious acquisition of a skill through practice. Rather, it stemmed from separate mental representations that were rearranged during sleep after training had taken place. First, although in the sleep group the times taken to solve each sequence of the string of digits decreased overnight in both ‘solvers’ (the 60% who gained insight later on) and ‘nonsolvers’ (who did not), this overnight decrease in reaction time was much smaller in the solvers than in the nonsolvers. Second, compared with the nonsolvers, the responses to the first digits in a sequence were delayed in the solvers as early as the end of the training session. It seems that the solvers spent time analysing the task during the training and retest sessions. Nevertheless, the solvers were the fastest to find the final solution in a sequence because they were aware of and applied the hidden rule.

The primitive elements of the task that the participants gleaned during training seem to have been reorganized during sleep, eventually leading them to become conscious of the hidden rule the following morning. Sleep has been implicated in learning and memory in the adult brain and is thought to favour the ‘off-line’ processing of new memories³. Wagner and colleagues’ data can be viewed as an extreme case of memory processing, in which the reorganization of primitive representations leads to a new conscious knowledge that entirely changes and improves the subject’s ability to crack the problem.

This study, of course, raises plenty of questions. What are the neural correlates of the processing of the primitive representations during sleep that lead to the gain of insight the following day? During wakefulness, learning of the hidden rule is known to be related to activation of two parts of the brain in particular — the perirhinal cortex and superior parietal lobule². Do these areas participate in the off-line processing during sleep? What are the neuronal interactions, reinforced during sleep, that underpin the emergence of the conscious knowledge?

There is also the issue of whether all stages of sleep participate in these processes — that is, what does the ‘sleep committee’ that

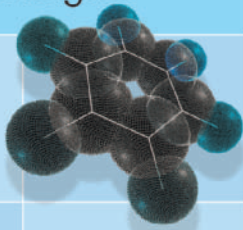
Scientific insights		Artistic creativity	
Friedrich Kekulé Ring-like structure of benzene ▶		Robert Louis Stevenson Key scenes in the novel <i>The Strange Case of Dr Jekyll and Mr Hyde</i> ▶	
 Otto Loewi Principle of chemical neurotransmission		 Samuel Taylor Coleridge The epic poem <i>Kubla Khan</i>	
Elias Howe The sewing machine ▶		Giuseppe Tartini The violin sonata <i>Il trillo del Diavolo</i> (Devil's Trill) ▶	
Herman Hilprecht Translation of cuneiform script on the 'stone of Nebuchadnezzar'	Dmitri Mendeleev The periodic table	Arthur Benson The poem <i>The Phoenix</i>	Jules Massenet Several operatic compositions

Figure 1 Instances of insight and creativity said to have followed sleep (or rather, except in the case of Massenet, occurring during or following a dream). Some may be apocryphal. But these examples foster the general feeling, tested in a controlled manner by Wagner *et al.*¹, that sleep may aid insight.

stimulates insight consist of? Explicit memory tasks are generally thought to be more sensitive to deprivation of slow-wave sleep⁴. On the other hand, most of the famous cases of scientific insight and artistic creativity (Fig. 1) are reported as emerging from a dream, which is a mental activity that occurs more frequently during the sleep pattern known as rapid-eye-movement (REM) sleep. Another possibility is that both non-REM and REM sleep are sequentially needed to optimize the memory^{5,6}. Wagner *et al.* give us the tools to explore these questions experimentally. Their paper constitutes steps forward in investigating the unpredictable and elusive phenomenon of insight, and in broadening the scope of the research on sleep, cognition and brain plasticity (the brain's ability to persistently change its structure and/or function).

The role that sleep plays in human creativity will be a mystery for some time yet. But at the very least, Wagner *et al.* give us good reason to fully respect our periods of sleep — especially given the current trend to recklessly curtail them. ■

Pierre Maquet and Perrine Ruby are at the Cyclotron Research Centre, University of Liège, Allée du 6 Août, 8 Sart Tilman, B-4000 Liège, Belgium.
e-mail: pmaquet@ulg.ac.be

1. Wagner, U., Gais, S., Haider, H., Verleger, R. & Born, J. *Nature* **427**, 352–355 (2004).
2. Rose, M., Haider, H., Weiller, C. & Buchel, C. *Neuron* **36**, 1221–1231 (2002).
3. Maquet, P., Smith, C. & Stickgold, R. (eds) *Sleep and Brain Plasticity* (Oxford Univ. Press, New York, 2003).
4. Plihal, W. & Born, J. *Psychophysiology* **36**, 571–582 (1999).
5. Stickgold, R., Whidbee, D., Schirmer, B., Patel, V. & Hobson, J. A. *J. Cogn. Neurosci.* **12**, 246–254 (2000).
6. Gais, S., Plihal, W., Wagner, U. & Born, J. *Nature Neurosci.* **3**, 1335–1339 (2000).

Biogeochemistry

Carbon budget in the black

Michael W. I. Schmidt

A significant fraction of a common organic component of marine sediments has an unexpected source, providing a fresh context for studies of the global carbon cycle in oceanic and terrestrial settings.

The global carbon cycle is both a consequence and a determinant of the state of the planet. Not surprisingly, then, a small army of scientists is at work studying the exchange of various guises of the element between land, atmosphere and ocean. One of those guises is 'black carbon', a relatively inert form that is produced by incomplete combustion during wildfires or the industrial burning of fossil fuels. It initially accumulates in soils, and is then distributed to lake and river sediments, and eventually to marine sediments.

On page 336 of this issue, Dickens *et al.*¹ describe the results of their investigations into how much black carbon is stored in marine sediments. They have analysed sediment laid down in pre-industrial times, and conclude that the fire-derived carbon content has been overestimated. As much as half of the organic carbon previously identified as black carbon instead appears to be fossil 'graphitic' black carbon — that is, carbon derived from the weathering of rocks.

This conclusion is based on two pillars of evidence. One is that although it was previously not possible to distinguish chemically between fire- and rock-derived carbon, Dickens *et al.* have done so using a novel chemical approach. They have therefore been able to isolate graphitic black carbon. The other arises from application of the isotope ¹⁴C as an analytical tool. Organic carbon that originates in terrestrial wildfires has a ¹⁴C signature. Carbon derived from fossil-fuel burning in the industrial era, or

from geologically much older sources (such as metamorphic rocks), has no such signature — that is, it is 'radiocarbon dead'. The literature on this topic is rather scant. But by combining their own data from

pre-industrial samples with previous results, Dickens *et al.* make the case that the proportion of fire-derived black carbon in global marine sediments has been overestimated.

Here, however, a comment by Alexandre Dumas in the nineteenth century comes to mind: "All generalizations are dangerous, even this one." The results of Dickens *et al.* stem from a limited number of samples, originating mostly from the continental shelf off the American Pacific northwest. They may therefore be representative of such regions — tectonically active continental margins, with high erosion rates on land and high sedimentation rates in the adjacent marine environment. But there are not enough data from other places to be sure that the results apply more widely to the same extent. For example, sediments from open ocean areas are represented by only one sample from the central Pacific.

As always, more data would be helpful, and would increase confidence in Dickens and colleagues' conclusions. But to be fair, Dumas' point applies to much research in this field, and the authors do present a plausible case. If their results are indeed typical for many regions of the world, they will have considerable influence on both oceanographic and terrestrial studies of the carbon cycle.

The authors themselves describe some of the implications. For instance, they consider that radiocarbon dating of sedimentary organic carbon may need a rethink. And they calculate that 7×10^{11} g of graphitic black



Figure 1 Carbonized conundrum. Black carbon is formed during wildfires in ecosystems such as savannah grasslands and high-latitude coniferous forests. a, Immediately after a fire, the local area is covered by large pieces and dust-sized particles of black carbon. b, After 60 years, a few charred tree stumps remain among a new generation of trees, and only small quantities of black carbon can be detected in the soil. c, Its fate could be export by rivers into the oceans (the Yenisei River, which runs into the Arctic Ocean, is shown here). But given the finding¹ that marine sediments contain less fire-derived black carbon than previously thought, some of the carbon must be degraded at an earlier stage or stored elsewhere.

M. W. I. SCHMIDT

Virology

A class act

Theodore S. Jardetzky and Robert A. Lamb

Membrane fusion occurs in many situations in living organisms — when certain viruses enter host cells, for instance. Three crystal structures shed light on the protein rearrangements that bring about such fusion.

carbon is deposited annually in the world's oceans. This is only some 0.5% of the total amount of organic carbon estimated to be buried in marine sediments each year. But because graphitic black carbon is thought to be created and destroyed very slowly, Dickens *et al.* think that this loop may act as a significant 'holding pool' of organic carbon distinct from the main parts of the carbon cycle.

As far as terrestrial studies are concerned, a host of questions arise. Black carbon is produced in enormous quantities by wildfires and by burning fossil fuels², and it is virtually ubiquitous in the terrestrial environment³. But if much less black carbon is ending up in marine sediments than we had thought, where is the excess? Possibilities include retention in the marine water column, or in rivers, soils or the atmosphere. If it is not stored in these 'sinks', it might be degraded before it reaches them, possibly both mechanically and chemically.

There has, however, been little unambiguous evidence that either process is occurring^{3–5}. As an example, take high-latitude coniferous forests, which are large, fire-prone ecosystems (Fig. 1). These forests burn quite often, but the amounts of black carbon on the forest floor and underlying soil can be surprisingly small⁶. Is the black carbon re-burnt during subsequent fires? Do freezing and thawing break it up into smaller particles, thus making it more susceptible to chemical degradation? There is evidence that atmospheric black carbon can be oxidized and become water-soluble⁷, a fate to which terrestrial black carbon might also succumb. Further, water-soluble carbon, discharged in groundwater and rivers, may not be accounted for in carbon budgets at the regional scale⁸.

The work by Dickens *et al.*¹ gives us a lot to think about. If the quantity of black carbon in marine sediments has indeed been overestimated, then the rates at which it is degraded have been underestimated and its contribution to the global carbon cycle^{2,9} must be redefined. The terrestrial aspect of that cycle is especially poorly understood; black carbon is just one of the many pieces still missing from the jigsaw¹⁰.

Michael W. I. Schmidt is in the Department of Geography, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.
e-mail: michael.schmidt@geo.unizh.ch

- Dickens, A. F., Gélina, Y., Masiello, C. A., Wakeham, S. & Hedges, J. I. *Nature* **427**, 336–339 (2004).
- Kuhlbusch, T. A. J. *Science* **280**, 1903–1904 (1998).
- Schmidt, M. W. I. & Noack, A. G. *Glob. Biogeochem. Cycles* **14**, 777–794 (2000).
- Shneour, E. *Science* **155**, 991–992 (1966).
- Baldock, J. A. & Smernik, R. J. *Org. Geochem.* **33**, 1093–1109 (2002).
- Czimczik, C. I. *et al.* *Glob. Biogeochem. Cycles* **17**, 1020 (2003).
- Decesari, S. *et al.* *Atmos. Environ.* **36**, 1827–1832 (2002).
- Siemens, J. *Science* **302**, 1681 (2003).
- Seiler, W. & Crutzen, P. J. *Clim. Change* **2**, 207–247 (1980).
- Prentice, I. C. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 881 (Cambridge Univ. Press, 2001).

Many viruses, such as the dengue fever, influenza and human immunodeficiency viruses, are encased in a lipid membrane. To reproduce, such 'enveloped' viruses must enter a host cell by fusing their own membrane coat with that of the cell. Fusion is caused by specific proteins in the viral membrane, and at least two classes of these proteins have been identified. In both classes, tightly regulated conformational changes are involved in membrane fusion. But class I and class II fusion proteins have distinct structural features — raising the expectation that the details of the two fusion mechanisms would also be unique. Yet the crystal structures of the post-fusion forms of three class II proteins, described elsewhere in this issue^{1,2} and in *EMBO Journal*³, suggest an unanticipated mechanistic parallel between class I and class II fusion proteins. The new structures also provide the first view of a key region of the proteins (the fusion loop) in the conformation that is predicted to insert into the target membrane.

The haemagglutinin protein from influenza virus has provided the model for class I fusion machines, because the atomic structures of three different forms of this protein have been determined⁴. The class I model, in

its essential features, applies to many unrelated virus families and their fusion proteins, including the HIV gp120 protein, the F proteins from paramyxoviruses, the retroviral SU/TM proteins and the Ebola virus Gp2 protein. For influenza haemagglutinin, a drastic refolding of the protein is observed between its pre-fusion and post-fusion forms (Fig. 1). Related refolding events have been inferred for the other class I proteins⁵.

The second class of viral fusion protein was postulated⁶ from the striking structural conservation of the pre-fusion forms of three particular proteins. Two of these, the E proteins of tick-borne encephalitis virus (TBE)⁷ and dengue virus⁸, are likely to be representative of a large and diverse family of RNA viruses, the Flaviridae, which also includes yellow fever virus, West Nile virus and hepatitis C virus. The third protein — the E1 protein of Semliki Forest virus⁶ — represents a family of RNA viruses called Togaviridae, of which the best known is the rubella virus, which causes German measles. The post-fusion structures of dengue virus E protein¹, Semliki Forest virus E1 protein² and TBE E protein³ have now been determined; they reveal a marked and surprising convergence of the class I and II fusion mechanisms.

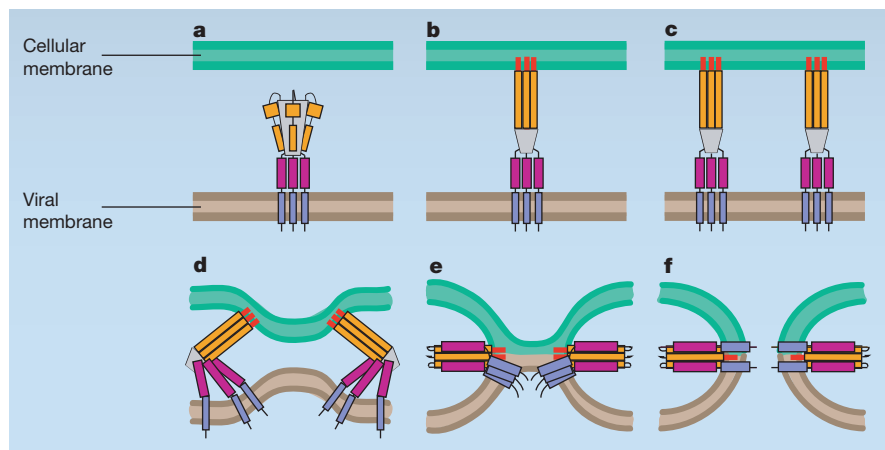


Figure 1 Proposed mechanism for membrane fusion by class I fusion proteins. **a**, The metastable conformation of a trimeric generalized fusion protein, with helical domain A in orange, helical domain B in pink, and the transmembrane domain in purple. **b**, After binding to a receptor on the cellular membrane, or on exposure to the low pH found in intracellular compartments (endosomes), the protein forms an extended conformation and the hydrophobic fusion peptide (red) inserts into the target membrane. **c**, Several trimers are thought to be involved. **d**, Protein refolding begins. The free energy thereby released causes the membranes to bend towards each other. **e**, Formation of a restricted hemifusion stalk allows the lipids in the outer leaflets of the membranes to mix. **f**, Protein refolding completes, forming the final, most stable form of the fusion protein, with the fusion peptide and transmembrane domain anti-parallel to each other but in the same membrane. Only **a** and **f** have been observed by crystallography, but biochemical data support many of the proposed steps.

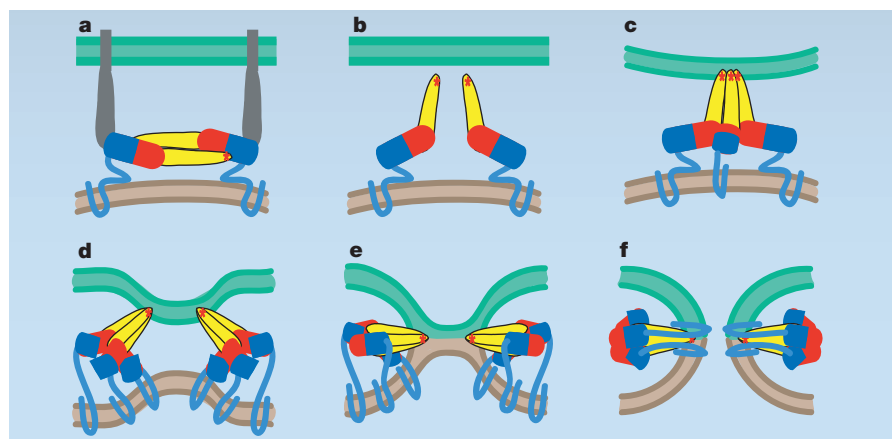


Figure 2 Proposed mechanism for fusion by class II proteins, incorporating the new results^{1–3}. **a**, The dimeric E protein binds to a cellular receptor (grey) and the virus is internalized to endosomes. Membrane fusion, which will release the virus into the body of the cell, takes place within endosomes. Domain I is in red, domain II in yellow, and domain III in light and dark blue (the lighter blue shows a predicted but unsolved structure). **b**, The acidic pH inside endosomes causes domain II to swing upward, permitting E monomers to rearrange laterally. **c**, The fusion loop (red dot) inserts into the outer leaflet of the host-cell membrane, enabling trimer formation. **d**, The formation of trimer contacts extends from the top to the bottom of the molecule. Domain III shifts and rotates to create contacts, bending the membrane. **e**, The formation of further contacts leads to unrestricted hemifusion. **f**, The final most stable form of the protein. Only **a** and **f** have been observed by crystallography but biochemical data support many of the other steps.

To understand the details of these findings, we need a little more background. Class I fusion proteins are composed of three identical protein subunits, the functional forms of which are generated from a precursor that is cleaved into two pieces. The carboxy-terminal end of one piece is anchored to the viral membrane; the other end (the new amino terminus) has a characteristic stretch of 20 hydrophobic amino acids — the fusion peptide. During fusion, the three fusion peptides become exposed and are inserted into the target cell membrane (Fig. 1), generating a differently folded intermediate that is anchored to both cellular and viral membranes. This intermediate can be inhibited by short peptides from regions of the protein that form the final core of the post-fusion form: a new HIV-inhibiting peptide, Fuzeon (enfuvirtide), blocks HIV infection at this step.

In the absence of inhibitors, however, the protein further refolds into a trimeric hairpin shape, a process that is tightly linked to fusion. An essential feature of class I proteins is the formation at this time of a trimeric helical coiled-coil rod adjacent to the fusion peptide. This structure may act as a template for the refolding of protein segments — often helical — that are near to the segments anchored in the virus membrane. Refolding relocates the fusion peptides and transmembrane anchor domains to the same end of the coiled coil, bringing viral and cellular membranes together to catalyse fusion^{4,5,9,10}.

Class II fusion proteins have distinctly different structural features: they are predominantly non-helical, instead having a β -sheet-type structure; they are not cleaved during biosynthesis; and the portion that

inserts into the target membrane is thought to be an internal hydrophobic fusion loop. The proteins have a three-domain architecture, in which domain I begins at the amino terminus, domain II contains the internal fusion loop, and domain III is at the carboxy terminus (Fig. 2). In cryo-electron-microscopy reconstructions of virus particles, dimers of E or E1 proteins lie flat upon the viral surface¹¹. At pH values equivalent to those in endosomes (the cellular compartments in which internalized viruses meet an acidic pH), the proteins assemble irreversibly into trimers. But for the truncated, soluble proteins this occurs only in the presence of membrane lipids, suggesting that interactions with the membrane are important.

The crystal structures reported by Modis *et al.*¹, Gibbons *et al.*² and Bressanelli *et al.*³ clarify the class II fusion mechanism. Another cryo-electron-microscopy study¹² had suggested that a large portion of β -sheet structure might insert into the membrane during fusion. But the new structures do not support this model: they show significant reorganization of the orientations of the relative domains, with relatively limited protein refolding and only the fusion loop inserting into the membrane. The reorientations are accompanied by disassembly of the dimer and formation of an extensive interface in the trimer that involves all three domains. This would result in the trimer lifting up from the virus surface and projecting the internal fusion loop towards the target membrane (Fig. 2). It seems most likely from the location of hydrophilic residues that the fusion loop inserts only into the outer half (leaflet) of the target membrane.

So far this is all very different from influenza haemagglutinin (and, by inference, from other class I proteins), in which refolding, rather than reorganization, is the key. But a closer look at the trimeric forms of the E and E1 proteins reveals some unexpected similarities to class I proteins. In these forms, domains I and II produce an extended, thick, rod-like structure, with the fusion loops positioned at the tips — analogous to the coiled-coil rods seen in class I proteins. Moreover, the displacement of domain III, and its relocation along the outside of the rod, produces a familiar hairpin-like structure. In the class II structures, an extended groove along the outside of the trimer is poised to interact with the segments connecting to the anchor domains in the viral membrane — again reminiscent of class I proteins. Overall, the new structures suggest that trimerization of class II proteins assembles a trio of hairpins that juxtapose fusion loops and anchor domains at the same end of a roughly 100-Å-long structure. The similarities to the class I pathway hint at evolutionary convergence upon a common mechanistic approach, but with unrelated protein architectures.

These structures^{1–3} open new approaches to the development of inhibitors of class II fusion proteins, analogous to existing class I inhibitors. The structures also — together with the pre-fusion conformations and electron-microscopic reconstructions — provide new insights into the assembly and regulation of class II fusion machines. For proteins from both classes, the transition to the post-fusion state is associated with an irreversible conformational change that probably provides the energy for membrane fusion. But assembly of the pre-fusion structures into metastable ‘cocked’ states that can be triggered appropriately is achieved very differently. The new structures suggest that, for class II proteins, target membranes might have a crucial catalytic or facilitating role in lowering the barrier to transition, presumably by orienting protein monomers at the membrane surface and so promoting the necessary rearrangements. ■

Theodore S. Jardetzky and Robert A. Lamb are in the Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208-3500, USA.

e-mails: tedj@northwestern.edu

ralamb@northwestern.edu

- Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. *Nature* **427**, 313–319 (2004).
- Gibbons, D. L. *et al.* *Nature* **427**, 320–325 (2004).
- Bressanelli, S. *et al.* *EMBO J.* (in the press).
- Skehel, J. J. & Wiley, D. C. *Annu. Rev. Biochem.* **69**, 531–569 (2000).
- Colman, P. M. & Lawrence, M. C. *Nature Rev. Mol. Cell Biol.* **4**, 309–319 (2003).
- Lescar, J. *et al.* *Cell* **105**, 137–148 (2001).
- Rey, F. A. *et al.* *Nature* **375**, 291–298 (1995).
- Modis, Y. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 6986–6991 (2003).
- Melikyan, G. B. *et al.* *J. Cell Biol.* **151**, 413–423 (2000).
- Russell, C. J., Jardetzky, T. S. & Lamb, R. A. *EMBO J.* **20**, 4024–4034 (2001).
- Zhang, W. Z. *et al.* *Nature Struct. Biol.* **10**, 907–912 (2003).
- Gibbons, D. L. *et al.* *Cell* **114**, 573–584 (2003).

Atmospheric chemistry

Getting fresh at the weekend

Atmos. Chem. Phys. **3**, 2225–2232 (2003)

Nitrogen oxides in the atmosphere are linked to a suite of adverse effects on both climate and human health. Some 50% of this pollution is thought to come from the burning of fossil fuels. As nitrogen oxides survive for only about a day in the atmosphere, their levels would be expected to dip at the weekend as human activity diminishes. Satellite measurements now show that this is indeed the case throughout the industrialized world.

S. Beirle and colleagues used data from the European Space Agency's ERS-2 satellite to track levels of NO₂ in the atmosphere's lowest layer, the troposphere. Their analysis covered the United States, Europe, the Middle East and East Asia.

In regions with a Christian tradition, levels of NO₂ plummeted on Sundays: Milan, in Italy, experienced a drop of 60% relative to weekday levels. Arab cities showed lowest NO₂ levels on Friday, the Islamic day of rest, whereas Jerusalem was least polluted on the Jewish sabbath, Saturday.

China did not show a weekend effect, despite having a Monday–Friday working week. This may be because China's nitrogen oxide emissions are largely due to power plants and industry, which operate throughout the week, whereas in the West emissions from motor vehicles play an important role.

Michael Hopkin

Cardiovascular biology

Division and development

J. Cell Biol. **164**, 97–109 (2004)

The shape of a heart conjures up thoughts of roses, rings and romance — but little is known about how the pumping human version acquires its characteristic shape. Sigolène M. Meilhac *et al.* have found evidence that the curves arise through carefully arranged cell growth.

The researchers studied genetically labelled single precursor cells in a mouse embryo's growing heart, as it transformed from a primitive cardiac tube into a four-chambered organ. The label was inherited only by daughter cells, allowing the team to examine where they ended up.

The authors find that, rather than daughter cells spreading randomly around their mother, they take up specific locations in each region. In the elongated outflow tract, for example, the cells arrange themselves in roughly vertical lines. To form the rounded left ventricle that bulges out from the primitive tube, daughter cells

position themselves like the spokes of a wheel.

The finding contradicts a previous idea that organs gain their shape via faster or slower rates of cell division and death in different regions. Although it remains unclear what controls the oriented cell growth, the group suggests that other organs, such as the liver or kidneys, may also use the mechanism to achieve their perfect form.

Helen Pearson

Chemistry

Lizards line up

J. Am. Chem. Soc. doi:10.1021/ja039515r (2004)

Building porous silicates from individual molecules is a useful way to create nanoscale channels that are bristling with organic functional groups. This method is often preferable to starting with grains of silica and then forcing organic molecules into existing pores, which usually results in much lower loading densities.

Qingmin Zhang and colleagues have developed a template molecule that can be assembled into a densely functionalized porous material. They compare this template to a lizard. First, the lizards' heads join up to make the basic silicate structure. The lizards' long tails take up enough space to ensure that nanoscale channels run through the newly formed silicate scaffold. A simple hydrolysis reaction cuts the lizards' tails off and washes them out of the pores, leaving their bodies covering the inner surface of the channels. In this case, the lizards' bodies are alanine residues, but the authors claim that this method could be used for a wide variety of functional groups.

They also speculate that this may be a swift route to incorporating asymmetric molecules on the internal surfaces of porous materials, for use as catalysts or in sensing devices.

Mark Peplow

Pharmacology

Herbal medicine at work

J. Clin. Invest. **113**, 137–143 (2004)

The herbal tea Yin Zhi Huang, a concentrated extract of four different plants, has been used in Asia for centuries to prevent and treat jaundice in newborn babies. The treatment helps to remove excess bilirubin — a breakdown product of worn-out red blood cells — from the liver. Wendong Huang *et al.* have discovered how the remedy may work.

The team used mice to study a receptor protein called CAR that is thought to aid bilirubin clearance. Treatment with Yin Zhi Huang increased the rate of bilirubin removal in mice that had the protein, and also induced the expression of various genes

involved in bilirubin metabolism. These effects were not seen in mice lacking CAR, suggesting that Yin Zhi Huang functions through this protein.

One particular ingredient from the herbal concoction was similarly effective: scoparone accelerated bilirubin removal in CAR-expressing mice, and activated the expression of genes that are targets of CAR in cultured liver cells. Huang *et al.* conclude that scoparone is an active component of Yin Zhi Huang, and suggest that drugs that target CAR could help to treat jaundice.

Helen R. Pilcher

Animal behaviour

Snap judgements

J. Exp. Biol. **207**, 393–398 (2004)

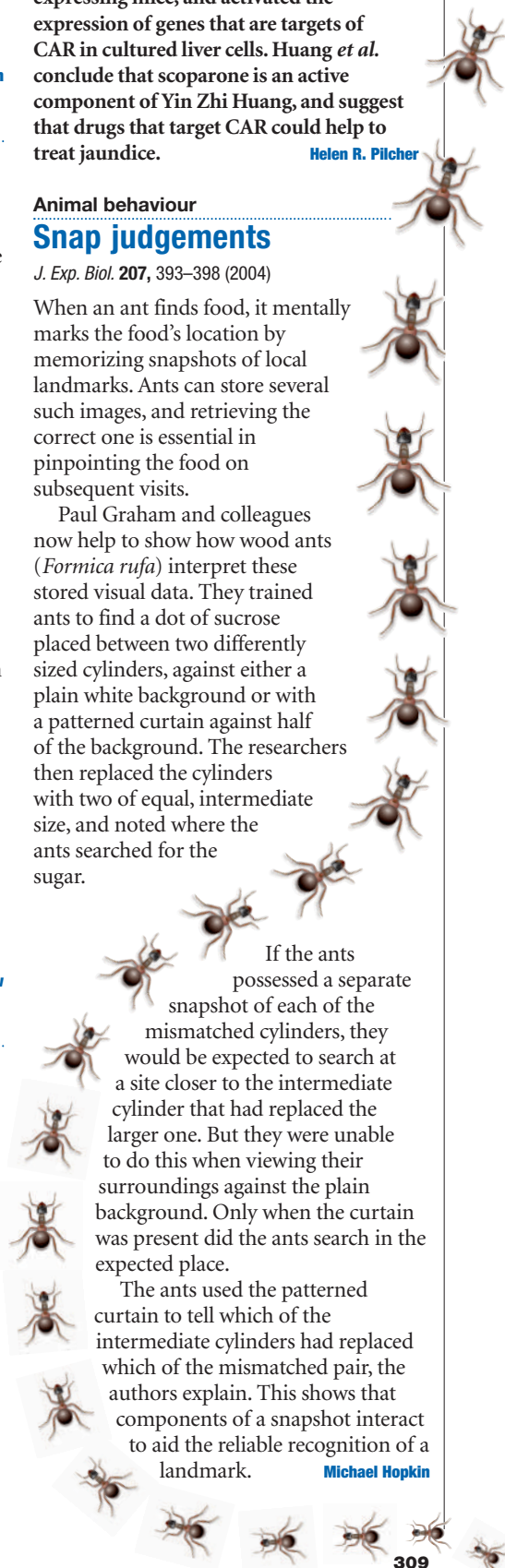
When an ant finds food, it mentally marks the food's location by memorizing snapshots of local landmarks. Ants can store several such images, and retrieving the correct one is essential in pinpointing the food on subsequent visits.

Paul Graham and colleagues now help to show how wood ants (*Formica rufa*) interpret these stored visual data. They trained ants to find a dot of sucrose placed between two differently sized cylinders, against either a plain white background or with a patterned curtain against half of the background. The researchers then replaced the cylinders with two of equal, intermediate size, and noted where the ants searched for the sugar.

If the ants possessed a separate snapshot of each of the mismatched cylinders, they would be expected to search at a site closer to the intermediate cylinder that had replaced the larger one. But they were unable to do this when viewing their surroundings against the plain background. Only when the curtain was present did the ants search in the expected place.

The ants used the patterned curtain to tell which of the intermediate cylinders had replaced which of the mismatched pair, the authors explain. This shows that components of a snapshot interact to aid the reliable recognition of a landmark.

Michael Hopkin



Changes in grey matter induced by training

Newly honed juggling skills show up as a transient feature on a brain-imaging scan.

Does the structure of an adult human brain alter in response to environmental demands^{1,2}? Here we use whole-brain magnetic-resonance imaging to visualize learning-induced plasticity in the brains of volunteers who have learned to juggle. We find that these individuals show a transient and selective structural change in brain areas that are associated with the processing and storage of complex visual motion. This discovery of a stimulus-dependent alteration in the brain's macroscopic structure contradicts the traditionally held view that cortical plasticity is associated with functional rather than anatomical changes.

Animal studies indicate that experience-related changes may occur in mammalian brain structures, but so far there has been no evidence of comparable alterations in the human brain^{3–5}. To investigate this possibility, we divided a homogeneous group of volunteers (21 female, 3 male; mean age, 22 yr $\pm$ 1.6 s.d.), who were matched for sex and age, into two groups, designated as jugglers and non-jugglers. Both groups were inexperienced in juggling at the time of their first brain scan.

Subjects in the juggler group were given 3 months to learn a classic three-ball cascade juggling routine. A second brain scan was

performed when they had become skilled performers (that is, when they could sustain juggling for at least 60 seconds). A third scan was carried out 3 months later; during the intervening period, none of the jugglers practised or attempted to extend their skills — for example, by learning a four-ball or a reverse cascade. In fact, most subjects were no longer fluent in three-ball cascade juggling by the time of the third scan.

We used voxel-based morphometry, a sophisticated objective whole-brain technique, to investigate subtle, region-specific changes in grey and white matter by averaging results across the volunteers. This method is based on high-resolution, three-dimensional magnetic-resonance imaging, registered in a common stereotactic space, and is designed to find significant regional differences by applying voxel-wise statistics in the context of gaussian random fields^{6,7}.

Group comparison at the beginning (the baseline) showed no significant regional differences in grey matter between jugglers and non-jugglers. In the longitudinal analysis, the juggler group demonstrated a significant (44 d.f., $P < 0.05$) transient bilateral expansion in grey matter in the mid-temporal area (hMT/V5) and in the left posterior intraparietal sulcus between the first and the



THE IMAGE BANK / GETTY

second scans. This expansion decreased in the third scan (Fig. 1). We found a close relationship in these regions between the transient structural grey-matter changes and the juggling performance. These findings were specific to the training stimulus, as the non-jugglers showed no change in grey matter over the same period.

Our results contradict the traditionally held view that the anatomical structure of the adult human brain does not alter, except for changes in morphology caused by ageing or pathological conditions. Our findings indicate that learning-induced cortical plasticity is also reflected at a structural level.

As all of our volunteers have normal fine-motor skills, we conclude that juggling, and consequently the perception and spatial anticipation of moving objects, is a stronger stimulus for structural plasticity in the visual areas (used for the retention of visual-motion information^{8,9}) than in the motor areas (involved in the planning and execution of coordinate motion — that is, the supplementary motor area and/or the motor cortex, cerebellum and basal ganglia).

Although the observed transient increase in grey matter takes place in specific motion-selective areas, the microscopic changes underlying these dynamic structural alterations are unclear. Macroscopic alterations may be based on changes at the level of

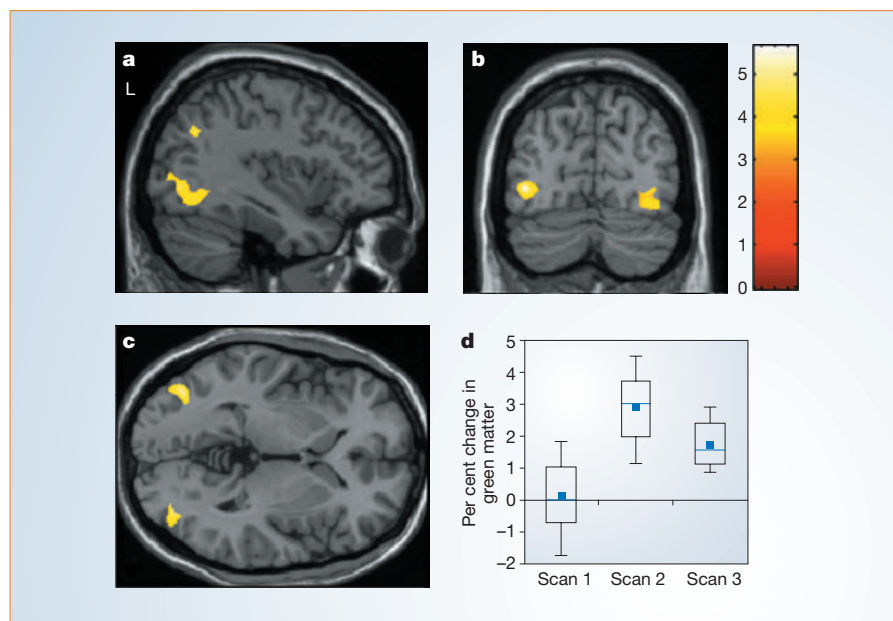


Figure 1 Transient changes in brain structure induced while learning to juggle. **a–c**, Statistical parametric maps showing the areas with transient structural changes in grey matter for the jugglers group compared with non-juggler controls. **a**, Sagittal view; **b**, coronal view; **c**, axial view. The increase in grey matter is shown superimposed on a normalized T1 image. The left side (L) of the brain is indicated. A significant expansion in grey matter was found between the first and second scans in the mid-temporal area (hMT/V5) bilaterally (left: $x, -43; y, -75; z, -2$, with $Z = 4.70$; right: $x, 33; y, -82; z, -4$, with $Z = 4.09$) and in the left posterior intraparietal sulcus ($x, -40; y, -66; z, 43$ with $Z = 4.57$), which had decreased by the time of the third scan. Colour scale indicates Z scores, which correlate with the significance of the change. **d**, Relative grey-matter change in the peak voxel in the left hMT for all jugglers over the three time points. The box plot shows the standard deviation, range and the mean for each time point.

synaptic bulk and neurites, or they might include increased cell genesis, for example, of glial or even neuronal cells⁴. Imaging results need to be compared with histological data for identification of the structural basis at the microscopic level of temporary, training-dependent structural changes in our brains.

Bogdan Draganski*, **Christian Gaser†**,
Volker Busch*, **Gerhard Schuierer‡**,
Ulrich Bogdahn*, **Arne May***

*Department of Neurology, and ‡Institute of
Neuroradiology, University of Regensburg,
Regensburg 93053, Germany

e-mail: arne.may@klinik.uni-regensburg.de

†Department of Psychiatry, University of Jena,
07740 Jena, Germany

1. Draganski, B. et al. *Nature Med.* **8**, 1186–1188 (2002).
2. Maguire, E. A. et al. *Proc. Natl Acad. Sci. USA* **97**, 4398–4403 (2000).
3. Kempermann, G., Gast, D. & Gage, F. H. *Ann. Neurol.* **52**, 135–143 (2002).
4. Trachtenberg, J. T. et al. *Nature* **420**, 788–794 (2002).
5. Grutzendler, J., Kasthuri, N. & Gan, W. B. *Nature* **420**, 812–816 (2002).
6. Ashburner, J. & Friston, K. J. *Neuroimage* **11**, 805–821 (2000).
7. Good, C. D. et al. *Neuroimage* **17**, 29–46 (2002).
8. Bisley, J. W. & Pasternak, T. *Cereb. Cortex* **10**, 1053–1065 (2000).
9. Sereno, M. I., Pitzalis, S. & Martinez, A. *Science* **294**, 1350–1354 (2001).

Competing financial interests: declared none.

Animal behaviour

Cognitive bias and affective state

Information processing by humans can be biased by their emotions — for example, anxious and depressed people tend to make negative judgements about events and to interpret ambiguous stimuli unfavourably^{1–4}. Here we show that such a ‘pessimistic’ response bias can also be measured in rats that are housed in unpredictable conditions^{5,6}. Our findings indicate that cognitive bias can be used as an indicator of affective state in animals, which should facilitate progress in animal-welfare studies.

We trained rats to respond by pressing a lever when they heard a tone associated with a positive event (delivery of a 45-mg food pellet) and to refrain from pressing the lever as a way to avoid a negative event (an unpleasant burst of white noise) when they heard another tone. Once the animals were able to score a correct response to each tone more than 50% of the time (binomial testing for three consecutive daily 30-min sessions), they were allocated to either ‘unpredictable’ housing, which induces symptoms of a mild depression-like state^{5,6}, or to ‘predictable’ housing.

In ‘unpredictable’ housing, between zero and two negative interventions were made at random times on any one day — for example, the cage might be unfamiliar or tilted, or it could contain a stranger of the same species; sometimes the light/dark cycle would be temporarily reversed or bedding

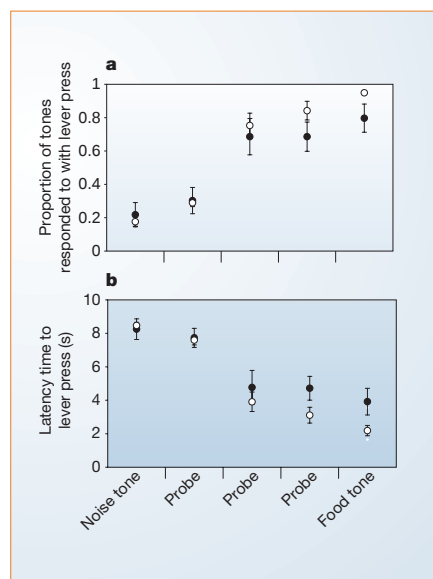


Figure 1 Mean (± 1 s.e.) responses to tones during 10 daily 30-min test sessions for male Lister hooded rats housed under ‘predictable’ (open circles, $n = 4$) and ‘unpredictable’ (filled circles, $n = 5$) conditions. **a**, Proportion of tones to which the animals responded to by pressing a lever. **b**, Latent time between sounding of the tone and pressing of the lever. ‘Noise’ and ‘food’ tones are the tones used during training (2 and 4 kHz, respectively, for about half of the rats, and 4 and 2 kHz, respectively, for the remaining rats). ‘Probe’ tones are non-reinforced, intermediate tones (2.5, 3, 3.5 kHz), each randomly interspersed with a probability of 0.085 between the reinforced training tones. Regression equations were calculated for each rat, correcting for nonlinear relationships by using binary logistic regression (for proportions) and logarithmic transformations for linear regression (for latencies). Animals were checked daily and remained healthy throughout the experiments.

left damp. These changes were never imposed simultaneously, and they were made at least two hours before or after test sessions. ‘Predictable’ housing, in contrast, was maintained as during training, with none of these interventions.

After nine days, during which training was continued, the rats were exposed to non-reinforced tones that had frequencies intermediate between those of the two food-delivery and noise-avoidance tones. Ten test sessions were held to investigate the animals’ anticipation of these positive or negative events, as judged by their lever-press response to these ambiguous tones.

The proportion of tones responded to by lever pressing (Fig. 1a) and the time taken to respond to the tones (mean response latencies; Fig. 1b) were calculated for each tone for each rat on each of the test days. Analysis of variance with repeated-measures (tone, test day) and a between-subjects factor (housing) revealed a housing $\times$ tone interaction ($F_{4,28} = 2.72$, $P < 0.05$).

The proportion of tones responded to with a lever press by individuals kept in unpredictable housing indicated that fewer lever presses were made in response to tones of frequency close to that of the food tone

(Fig. 1a) (in two-tailed t -tests, $t = 1.88$, d.f. = 7, $P = 0.1$). These rats were also slower to press the lever in response to the food tone and to the ambiguous tones that were close to it in frequency (Fig. 1b) ($t = -2.44$, d.f. = 7, $P < 0.05$). Both findings were still valid when only the responses by the rats to the ambiguous tones were analysed (proportions: $t = 1.92$, d.f. = 7, $P = 0.09$; latencies: $t = -2.42$, d.f. = 7, $P < 0.05$).

Overall, rats in unpredictable housing were slower to respond and tended to show fewer responses to ambiguous tones close to the positive tone and to this tone itself. The treatment groups did not differ ($P > 0.2$) in tests of feeding motivation (consumption speed of freely available food pellets⁷), anhedonia (amount of sucrose solution consumed^{5,6}), activity (hole-board test⁸), body-weight change across the test period, and response accuracy to training tones before and after the imposition of housing changes, indicating that none of these factors was likely to account for our findings.

By using ambiguous stimuli to probe animals’ relative anticipation of positive and negative events, we have shown that rats in unpredictable housing show behaviour indicating reduced anticipation of a positive event. This compares with findings for depressed or anxious humans, who also have reduced expectation of positive events^{1,4} and interpret ambiguous stimuli negatively³.

Our results call for further investigation of the underlying processes involved^{9,10}. We find no evidence of enhanced anticipation of the negative event. This may be due to a floor effect and could be revealed using, for example, lever-pressing and nose-poking as counterbalanced positive and negative responses. It is possible that our technique could be adapted to detect an enhanced expectation of positive events — a correlate of happy mood in humans⁴. Being able to assess positive as well as negative affect in animals is an important objective for animal welfare¹¹.

Emma J. Harding, Elizabeth S. Paul, Michael Mendl

Centre for Behavioural Biology, Department of
Clinical Veterinary Science, University of Bristol,
Langford House, Langford BS40 5DU, UK
e-mail: mike.mendl@bris.ac.uk

1. MacLeod, A. K. & Byrne, A. J. *Abnorm. Psychol.* **105**, 286–289 (1996).
2. Gotlib, I. H. & Krasnoperova, E. *Behav. Therapy* **29**, 603–617 (1998).
3. Eysenck, M. W. et al. *J. Abnorm. Psychol.* **100**, 144–150 (1991).
4. Wright, W. F. & Bower, G. H. *Organiz. Behav. Hum. Decis. Process* **62**, 276–291 (1992).
5. Willner, P. *Psychopharmacology* **134**, 319–329 (1997).
6. Zurita, A. et al. *Behav. Brain Res.* **117**, 163–171 (2000).
7. Nielsen, B. L. *Appl. Anim. Behav. Sci.* **63**, 79–91 (1999).
8. Fernandes, C. & File, S. E. *Pharmacol. Biochem. Behav.* **54**, 31–40 (1996).
9. Spruijt, B. M., van den Bos, R. & Pijlman, F. T. A. *Appl. Anim. Behav. Sci.* **72**, 145–171 (2001).
10. Berridge, K. C. & Robinson, T. E. *Trends Neurosci.* **26**, 507–513 (2003).
11. Dawkins, M. S. in *Coping with Challenge. Welfare in Animals including Humans* (ed. Broom, D.M.) 63–76 (Dahlem University Press, Berlin, 2001).

Competing financial interests: declared none.

Structure of the dengue virus envelope protein after membrane fusion

Yorgo Modis¹, Steven Ogata², David Clements² & Stephen C. Harrison¹

¹Howard Hughes Medical Institute, Children's Hospital and Harvard Medical School, 320 Longwood Avenue, Boston, Massachusetts 02115, USA

²Hawaii Biotech, Inc., 99–193 Aiea Heights Drive, Suite 200, Aiea, Hawaii, Hawaii 96701, USA

Dengue virus enters a host cell when the viral envelope glycoprotein, E, binds to a receptor and responds by conformational rearrangement to the reduced pH of an endosome. The conformational change induces fusion of viral and host-cell membranes. A three-dimensional structure of the soluble E ectodomain (sE) in its trimeric, postfusion state reveals striking differences from the dimeric, prefusion form. The elongated trimer bears three 'fusion loops' at one end, to insert into the host-cell membrane. Their structure allows us to model directly how these fusion loops interact with a lipid bilayer. The protein folds back on itself, directing its carboxy terminus towards the fusion loops. We propose a fusion mechanism driven by essentially irreversible conformational changes in E and facilitated by fusion-loop insertion into the outer bilayer leaflet. Specific features of the folded-back structure suggest strategies for inhibiting flavivirus entry.

Membrane fusion is the central molecular event during the entry of enveloped viruses into cells. The critical agents of this process are viral surface proteins, primed to facilitate bilayer fusion and triggered to do so by the conditions of viral interaction with the target cell. The best-studied example is the influenza virus haemagglutinin (HA), synthesized as a single-chain precursor and then cleaved into two chains, known as HA₁ and HA₂, during transport of the trimeric glycoprotein to the cell surface. The binding of HA₁ to a cell-surface receptor leads to endocytic uptake; acidification of the endosome triggers dramatic conformational rearrangement of HA₂. The latter is a two-stage process. Exposure of the amino-terminal 'fusion peptide' of HA₂ first allows it to insert into the endosomal membrane; subsequent folding-over of the entire HA₂ polypeptide chain brings together its N and C termini and thus forces the target-cell membrane (held by the fusion peptide) and the viral membrane (held by the C-terminal transmembrane anchor of HA) against each other.

HA is the prototype of a large class of viral fusion proteins—for example, those of other myxo- and paramyxoviruses such as measles virus, retroviruses such as HIV, and filoviruses such as Ebola virus (reviewed in ref. 1). All of these 'class I' viral fusion proteins are two-

chain products of a cleaved, single-chain precursor, and all bear a hydrophobic fusion peptide at or near the N terminus created by the cleavage². Moreover, in all class I fusion proteins, a three-chain, α -helical, coiled-coil assembles during the conformational change, drives the fusion peptide towards the target-cell membrane^{3–5}, and creates the central structural element of the fusion machinery^{6,7}.

An architecturally and evolutionarily distinct class of fusion proteins is found on flaviviruses, such as yellow fever, West Nile, and dengue viruses, and on alphaviruses, such as Semliki Forest and Sindbis viruses. These proteins associate with a second, 'protector' protein, the cleavage of which primes the fusion protein to respond to acidic pH. The fusion protein is called E in flaviviruses and E1 in alphaviruses. Structures have been determined for the ectodomains of three class II proteins in their prefusion states^{8–10}. Those of two flaviviruses, tick-borne encephalitis (TBE) and dengue viruses, are dimeric, both in solution¹¹ and on the viral membrane surface^{12,13} (Fig. 1). They have three domains, with folds based largely on β -sheets. One of these (domain II), an elongated, finger-like structure, bears a loop at its tip with a hydrophobic sequence conserved among all flaviviruses. Experiments with TBE virus show that this 'cd loop' (residues 98–109 in dengue type 2) is

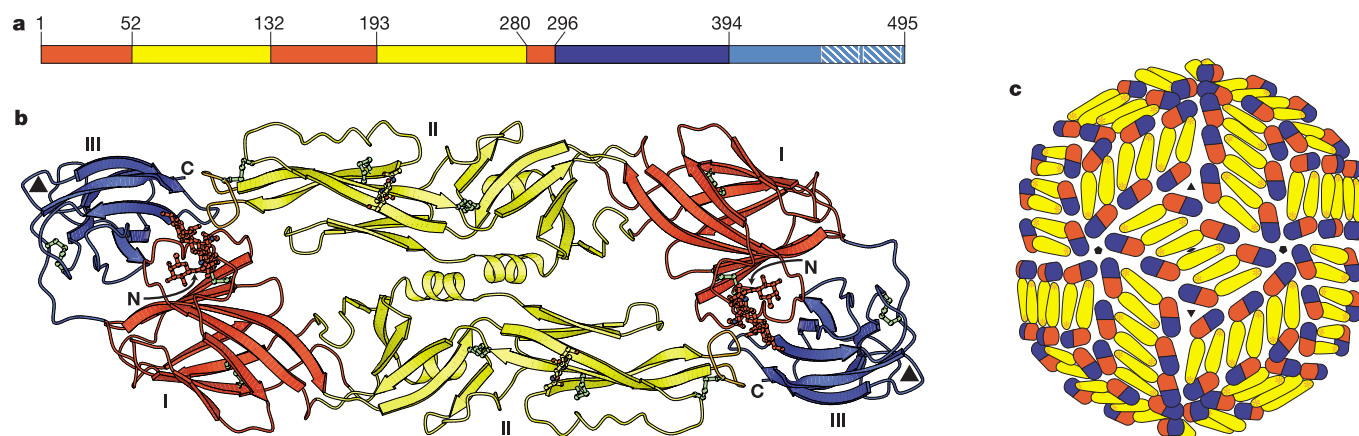


Figure 1 Structure of the dimer of dengue E soluble fragment (sE) in the mature virus particle. **a**, The three domains of dengue sE. Domain I is red, domain II is yellow, domain III is blue. A 53-residue 'stem' segment links the stably folded sE fragment with the C-terminal transmembrane anchor. **b**, The sE dimer¹⁰. This is the conformation of E in the

mature virus particle and in solution above the fusion pH. **c**, Packing of E on the surface of the virus. Electron cryomicroscopy image reconstructions show that 90 E dimers pack in an icosahedral lattice¹³.

responsible for attachment of soluble E ectodomains to target membranes and that the hydrophobic residues are essential for its activity¹⁴. These and other data, including results from studies of alphaviruses^{15,16}, which have a very similar cd-loop structure, support the view that the cd loop of class II fusion proteins has a function analogous to that of the N-terminal fusion peptide in class I fusion proteins: insertion into the host-cell membrane and provision of an attachment point for drawing host-cell and viral membranes together¹. We refer to the cd loop as the 'fusion loop', reserving 'fusion peptide' for the N-terminal segment of class I fusion proteins.

An important missing link in our picture of membrane fusion is a direct view of a fusion peptide or loop as it inserts into a target membrane. The crystal structure we describe here—of the soluble ectodomain of dengue virus type 2 E protein (sE) in its trimeric, postfusion conformation—provides such a link. The fusion loops of the three subunits come together to form a membrane-insertable, 'aromatic anchor' at the tip of the trimer. The fusion loop retains its prefusion conformation. Neighbouring hydrophilic groups restrict insertion to the proximal part of the outer lipid-bilayer leaflet. The entire ectodomain of the protein folds back on itself, directing the C-terminal, viral membrane anchor towards the fusion loop. Comparison with the prefusion structure of the same protein¹⁰ allows us to propose a mechanism for fusion driven by an essentially irreversible conformational change in the protein and assisted by membrane distortions imposed by fusion-loop insertion. Specific features of the folded-back structure suggest strategies for inhibiting flavivirus entry. The postfusion structure of an alphavirus fusion protein¹⁷ described in the accompanying paper shows an analogous conformation and leads to similar conclusions.

Membrane insertion and trimer formation

Like its TBE homologue, the dimer formed by dengue sE (residues 1–395 of E) dissociates reversibly^{10,11}. At acidic pH, dissociation is essentially complete at protein concentrations of 1 mg ml⁻¹; at neutral pH, the dissociation constant is one to two orders of magnitude smaller. The fusion loop at the tip of domain II would be exposed in the monomer^{8,10}, but exposure does not cause non-specific aggregation of the protein¹¹. Liposome coflotation experiments show that the fusion loop of monomeric TBE sE allows association with lipid membranes and that this membrane association catalyses irreversible formation of sE trimers at low pH¹⁸. Dengue E exhibits an identical behaviour: on acidification, sE dimers dissociate, bind liposomes and trimerize (see Methods). Membrane-associated sE is readily detected by electron microscopy of negatively stained preparations (Fig. 2a); chemical cross-linking confirms that the protein has trimerized (Fig. 2b). The trimers are tapered rods, about 70–80 Å long and 30–50 Å in diameter, with the long axis perpendicular to the membrane and their wide end distal to it. They tend to cluster on the liposome surface, often forming a continuous layer. These heavily decorated areas appear to have a greater than average membrane curvature, resulting in smaller vesicles (Fig. 2a). This observation suggests that E trimers can induce curvature, a property that may help promote fusion (see below). The dengue sE trimers can be solubilized with the detergent *n*-octyl- β -D-glucoside (β -OG); they remain trimeric at all pH values between 5 and 9, as determined by gel filtration chromatography (data not shown).

Domain rearrangements in the trimer

Crystals of the detergent-solubilized dengue sE trimers diffract to high resolution (2.0 Å); the structure was determined by molecular replacement (see Methods). The three domains of sE retain most of their folded structures, but undergo major rearrangements in their relative orientations, through flexion of the interdomain linkers (Fig. 3). Domain II rotates approximately 30° with respect to domain I, about a hinge near residue 191 and the kl hairpin

(residues 270–279), where mutations that affect the pH threshold of fusion are concentrated¹⁰. As a result of the rotation, the base of the kl hairpin is pulled apart, and the l strand forms a new set of hydrogen bonds with the D₀ strand of domain I, shifted by two residues from the hydrogen-bonding pattern in the dimer¹⁰. Although detergent is present, the kl hairpin does not adopt the open conformation seen in the dimer with bound β -OG¹⁰. The small hydrophobic core beneath this hairpin acts as a 'greased hinge' for the rotation between domains I and II.

Domain III undergoes the most significant displacement in the dimer-to-trimer transition. It rotates by about 70°, and its centre of mass shifts by 36 Å towards domain II. This folding-over brings the C terminus of domain III (residue 395) 39 Å closer to the fusion loop.

Changes in the secondary structure

The 10-residue linker between domains I and III accommodates their large relative displacement during trimer formation. The linker, which has a poorly ordered, extended structure in the prefusion dimer¹⁰ (Fig. 3a), inserts in as a short β -strand between strands A₀ and C₀ in domain I (Fig. 3b). As part of this rearrangement, the C-terminal region of A₀ peels away from C₀ and switches to the other β -sheet, thereby creating the surface for an annular trimer contact with the two other A₀ strands in the trimer.

The transition to the trimer state is irreversible. The refoldings just described may impart irreversibility by contributing a high barrier to initiation of trimerization (sE monomers do not trimerize at low pH without liposomes) and an even higher barrier to dissociation of trimers once they have formed.

Trimer contacts

Dengue E trimers assemble through both polar and nonpolar contacts in four areas: at the membrane-distal end of the trimer, at the base of domain II, at the tip of domain II, and at the packing interface between domains I and III (Fig. 4). The total surface buried per monomer during trimer assembly is 3,900 Å²—twice the 1,950 Å² per monomer buried in the dimer. An additional 1,035 Å² are buried within each monomer during the domain rearrangements observed in our structure. These numbers help to explain why trimers are much more stable than dimers in solution.

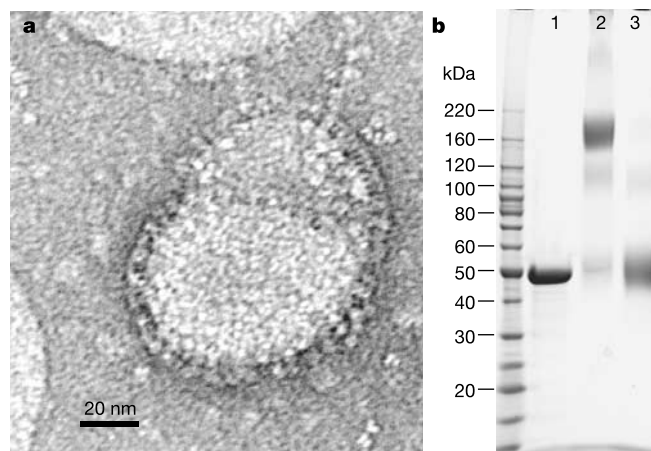


Figure 2 Trimer formation and membrane insertion of dengue E protein. **a**, Electron micrograph showing E trimers inserted into liposomes. The liposomes are heavily decorated with E trimers. A large portion of the trimer can be seen protruding from the membrane. The samples were stained with uranyl formate (see Methods). **b**, E trimers can be covalently cross-linked with ethylene glycol bis-succinimidyl-succinate (EGS) after insertion into liposomes. Lane 1, E solubilized from liposomes at pH 5.5 and not cross-linked. Lane 2, E solubilized from liposomes at pH 5.5 and cross-linked with EGS. Lane 3, E cross-linked with EGS at pH 7 in the absence of lipid.

They also help to account for the irreversibility of the fusion-activating conformational change. Additional trimer contacts are likely to be contributed by the stems, as described below.

An extended cavity, which runs along the three-fold axis, separates the trimer contact areas at the top and at the base of domain II. A narrow opening connects this cavity with the exterior solvent, but it may be occluded by the stem in the full-length protein (Fig. 4b). An anion, modelled as a chloride ion, lies on the three-fold axis near the tip of domain II. It is liganded by three amide nitrogens (from Lys 110 of each subunit) and three water molecules. Between this anion and the domain II tip, a small hydrophobic core underpins the nonpolar, bowl-shaped apex formed by the three clustered fusion loops (Fig. 4e).

The fusion loop

The fusion loops in the sE trimer have the same conformation as in the dimer¹⁰. Because the trimers are obtained by detergent extraction from liposomes, we conclude that this conformation is also present when the loop inserts into a membrane. Furthermore, because dimers can dissociate reversibly, the fusion loop is stable when fully exposed. It thus appears that the fusion loop retains essentially the same conformation, whether buried against another subunit, inserted into a lipid membrane, or exposed to aqueous solvent.

In the trimer, the three hydrophobic residues in the fusion loop conserved among all flaviviruses—Trp 101, Leu 107 and Phe 108—are fully exposed on the molecular surface, near the three-fold axis. They form a bowl-like concavity at the trimer tip, with a hydrophobic rim (Fig. 4e). There are no lipid or detergent molecules visible in the electron density near the fusion loop in either of our crystal forms (see Methods). Indeed, in the P321 crystal form, there can be no detergent micelle covering the fusion loop, as this region is involved in close crystal contacts with residues in domain III of a symmetry-related molecule. We conclude that detergent is required to dissolve away the liposome on which the trimer formed, and hence to solubilize the protein, but that once the protein has been extracted from the membrane, the three-fold-clustered fusion loops do not retain a tightly associated detergent micelle.

How deeply, then, do fusion loops penetrate into the membrane? Tryptophans tend to appear in membrane proteins at the interface between the hydrocarbon and head-group layers of the lipid¹⁹, but if the indole amine participates in a hydrogen bond, as is the case for Trp 101, the side chain may be completely buried in the hydrocarbon layer. We therefore propose that the E trimers penetrate about 6 Å into the hydrocarbon layer of the target membrane. They cannot penetrate further, because of exposed carbonyls and charged residues on the outside rim of the fusion-loop bowl (Fig. 4e). Thus, the fusion loop is held in the membrane mainly by an 'aromatic

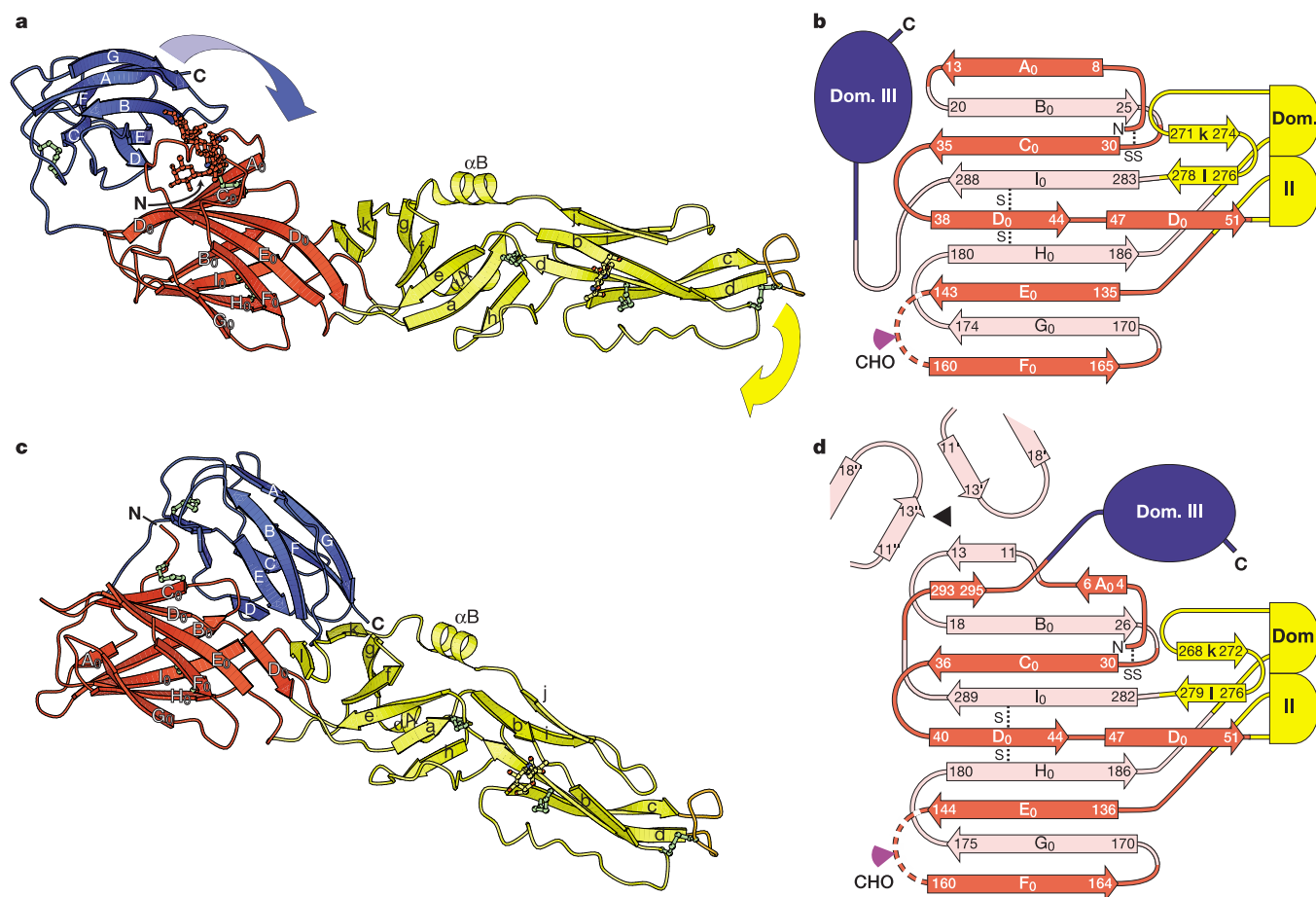


Figure 3 Domain rearrangements in the dengue sE monomer during the transition to trimer. **a**, An sE monomer in its prefusion conformation. This is the structure adopted in mature virus particles and in solution at pH > 7, when sE is a dimer. **b**, Schematic representation of the secondary structure of domain I and links to domains II and III in the prefusion conformation. **c**, An sE monomer in its postfusion conformation, as seen in sE trimers. The three domains have rotated and shifted with respect to each other, bringing

the C terminus 39 Å closer to the fusion loop (orange). The fusion loop retains essentially the same conformation before and after fusion. **d**, Secondary structure of domain I and its links to domains II and III in the trimeric, postfusion conformation. The domain I–III linker inserts between strands A₀ and C₀. The C-terminal region of A₀ flips out, switches to the other β-sheet, and creates an annular trimer contact with the two other A₀ strands in the trimer.

anchor' formed by Trp 101 and Phe 108. The bowl is lined by the hydrophobic side chains of Leu 107 and Phe 108, so that it cannot accommodate lipid headgroups. We expect that fatty-acid chains from the inner leaflet of the membrane may extend across to contact the base of the fusion-loop bowl, or that fatty-acid chains from the outer leaflet may bend over to fill it. In either case, insertion will produce a distortion in the bilayer, which could be important for the fusion process (see below).

A postfusion conformation

The folding back of domain III and the rearrangement of β -strands at the trimer interface projects the C terminus of sE towards the fusion loop, and positions it at the entrance of a channel, which extends towards the fusion loops along the intersubunit contact

between domains II (Fig. 4a, b). The 53-residue 'stem' connecting the end of the sE fragment with the viral transmembrane anchor could easily span the length of this channel, even if the stem were entirely α -helical. By binding in the channel, the stem would contribute additional trimer contacts with domain II of another subunit (Fig. 4b). The stem does indeed promote trimer assembly even in the absence of liposomes²⁰. In the virion, the stem forms two α -helical segments, which lie near the outer surface of the lipid bilayer and contact the subunit from which they emanate²¹. The proposed stem conformation places the viral transmembrane domain in the immediate vicinity of the fusion loop, just as in the postfusion conformations of class I viral fusion proteins. We therefore believe that the trimer we have crystallized represents a postfusion state of the protein.

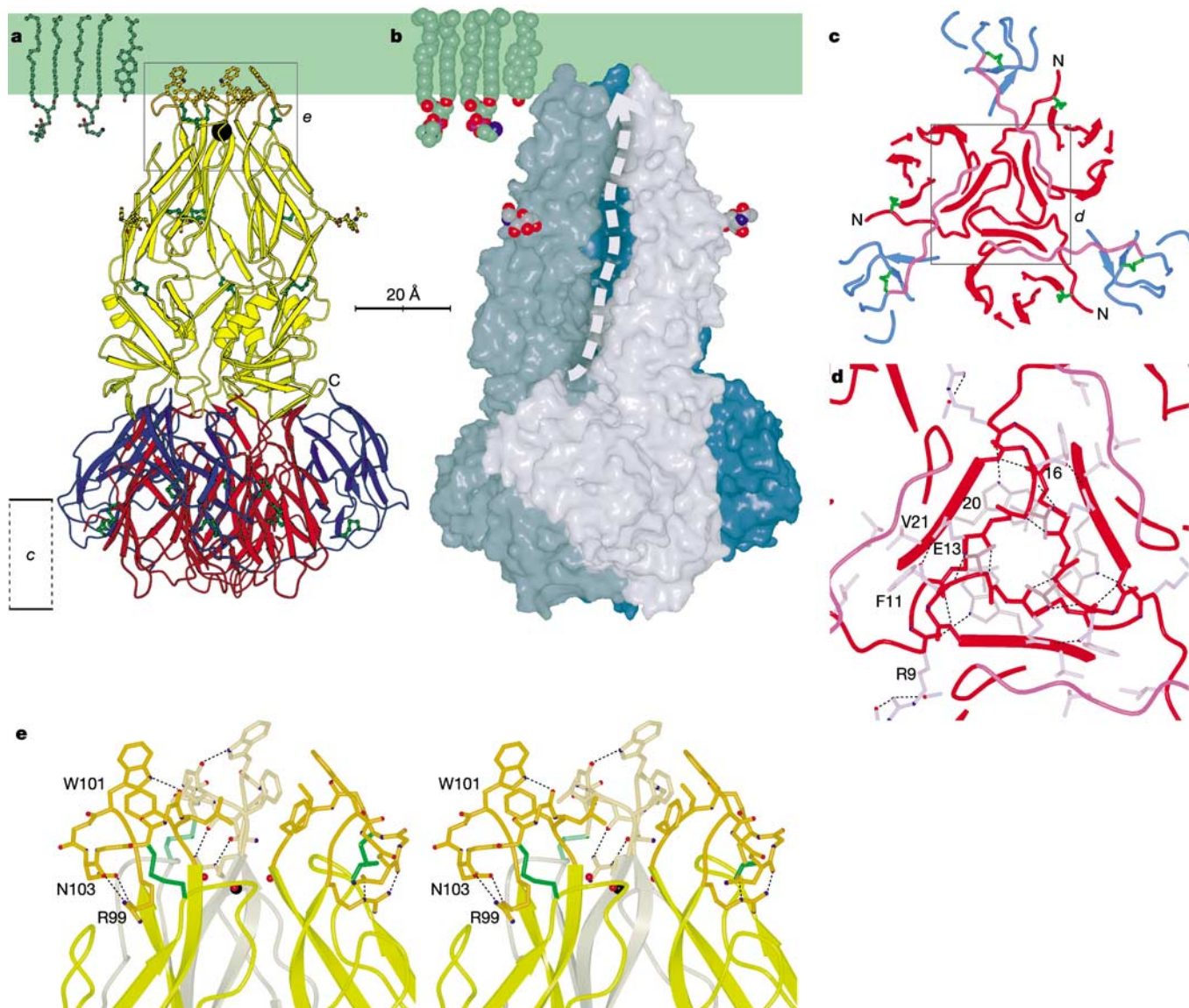


Figure 4 The dengue sE trimer. **a**, Ribbon diagram coloured as in Fig. 1b. Hydrophobic residues in the fusion loop (orange) are exposed. The expected position of the hydrocarbon layer of the fused membrane is shown in green. Representative lipids are shown to scale. A chloride ion (black sphere) binds near the fusion loop. **b**, Surface representation of the trimer. The dashed grey arrow indicates the most likely location for the stem (see text). An extended cavity is visible near the tip of the trimer; access to this cavity will probably be occluded by the stem. The glycan on Asn 67 and representative lipids are shown in space-filling representation. **c**, The membrane-distal end of the trimer,

where most trimer contacts are formed. The view is along the three-fold axis. **d**, Close-up of **c** showing trimer contacts. The A_0B_0 loop forms an annular trimer contact. The domain I–III linker (purple) adopts an extended conformation and forms additional trimer contacts. **e**, Close-up of the aromatic anchor formed by Trp 101 and Phe 108, in the fusion loop (orange). The three clustered fusion loops form a nonpolar, bowl-shaped apex, which is underpinned by a small hydrophobic core. Underneath, a chloride ion (black sphere) forms a trimer contact.

Mechanism of membrane fusion

The structure described here, combined with previous knowledge, allows us to propose the following mechanism for how conformational changes in the flavivirus E protein promote membrane fusion.

(1) E associates with a cell-surface receptor, probably through domain III^{22–28} (Fig. 5a); there is evidence for glycan-mediated interactions as well^{29–31}. Receptor binding leads to endosomal uptake.

(2) Reduced pH in the endosome causes E dimers on the virion surface to dissociate³², exposing their fusion loops and allowing domains I and II to flex relative to one another (Fig. 5b). Evidence for a pH-dependent hinge at the domain I–domain II interface includes the location of mutations that alter the pH threshold of fusion¹⁰, as well as the difference in orientation between the pre- and postfusion structures. Release of constraints imposed by dimer contacts may also allow the stem to extend away from the membrane. Some combination of these two sources of flexibility permits domain II to turn outward, away from the virion surface, and to insert its fusion loop into the target-cell membrane.

(3) Outward projection of domain II will destroy tight packing interactions on the virion outer surface, allowing lateral rearrangement of E monomers. Thus, the absence of trimer clustering in the virion¹³ is not, in principle, a barrier to trimer formation. Target membranes probably catalyse trimerization¹⁸, leading to a prefusion intermediate, in which the trimer bridges host-cell and viral membranes, with its fusion loops bound to the former and its transmembrane tail anchored in the latter (Fig. 5c). This species is analogous to the ‘prehairpin’ intermediate postulated for class I viral fusion mechanisms³³.

(4) Formation of trimer contacts spreads from the fusion loops at the trimer tip to domain I at the base. Domain III shifts and rotates, folding the C terminus of sE back towards the fusion loop (Fig. 5d). The length of the interdomain linker permits independent rotation of individual domains III, allowing for the spontaneous symmetry-breaking required at this point. Cooperativity and irreversibility occur only when the exchange of β -strands shown in Fig. 3d locks in

the final trimer interaction of domain I and the final folded-back position of domain III. Free energy released by this refolding can drive the two membranes to bend towards each other^{3–5}, forming apposing ‘nipples’³⁴ (Fig. 5d). Positive bilayer curvature induced by fusion-loop insertion might stabilize the lateral surfaces of such protrusions. A ring of trimers may be needed properly to deform the membrane. We cannot yet specify the number of trimers in such a ring nor how their conformational changes are coupled. It is possible that coupling is provided simply by the resistance of the membranes to deformation: only when several trimers act in concert can folding back reach the barrier of β -strand exchange.

(5) Formation of a ‘hemifusion stalk’, with proximal leaflets fused and distal leaflets unfused (Fig. 5e), is thought to be an essential intermediate in the membrane fusion reaction^{34–36}. Hemifusion could occur at any point during the conformational changes represented in Fig. 5d and e, depending on the length of the hemifusion stalk. We suggest that hemifusion would happen following the β -strand exchange step that locks domains III into their trimer positions. It must precede final zippering up of the stems, as full pore formation must occur before the transmembrane segments can reach their probable final positions around the periphery of the fusion loops. Hemifusion stalks can ‘flicker’ open into narrow fusion pores³⁵. Migration of the transmembrane segments along a transient pore will prevent its closing. Thus, if the transmembrane segments or adjacent stem regions ‘snap’ into place around the tips of domains II, formation of the symmetrical final structure (Fig. 5f) could drive the transition from stalk to pore.

(6) In the final state, the trimer has reached the conformation seen in our crystal structure, with the stems (not present in our current crystals) docked along the surface of domains II and with the fusion loops and transmembrane anchors now next to each other in the fused membrane (Fig. 5f).

Comparison with class I fusion

Class I and class II viral fusion proteins clearly have some common mechanistic features (Fig. 5). The most striking of these is a folding back of the protein during the fusion transitions, so that its two membrane attachment points come together in the postfusion structure. Class I proteins fold back by zippering up an ‘outer layer’ (at least partly α -helical) around a central, trimeric coiled-coil¹. Our structure of trimeric dengue sE shows that class II proteins do so by nucleating trimer formation around an elongated, finger-like fusion domain, by rearranging two other domains, and (probably) by zippering an extended C-terminal stem along the trimer surface.

Class II viral fusion proteins form trimers from monomers (dissociated homodimers in the case of flaviviruses; dissociated heterodimers in the case of alphaviruses³⁷), while class I proteins are trimeric in their prefusion state². But comparison of the pre- and postfusion states of influenza haemagglutinin—the only previous case where both structures are known for the same protein—shows that most of the trimer contacts in the latter state are not present in the former⁶. That is, just as in the trimerization of dengue E, the important trimer interactions in the final state form during the transition. These contacts are close to the three-fold axis, and they must be present before zippering up of an outer layer can occur. Indeed, the postulated prefusion intermediate is, both for class I fusion proteins and now for class II, a structure in which these central trimer contacts have formed but the zippering-up of the outer layer has not yet begun (Fig. 5).

Is our structure for the membrane-inserted state of the flavivirus fusion loops relevant also for class I fusion-peptide insertion? An NMR structure of an isolated, 20-residue influenza virus A fusion peptide associated with a detergent micelle suggests a slightly kinked α -helix, with its N and C termini embedded in the outer leaflet and the kink (at about residue 10) on the surface³⁸. Unlike the flavivirus and alphavirus fusion loops, however, the class-I fusion peptides

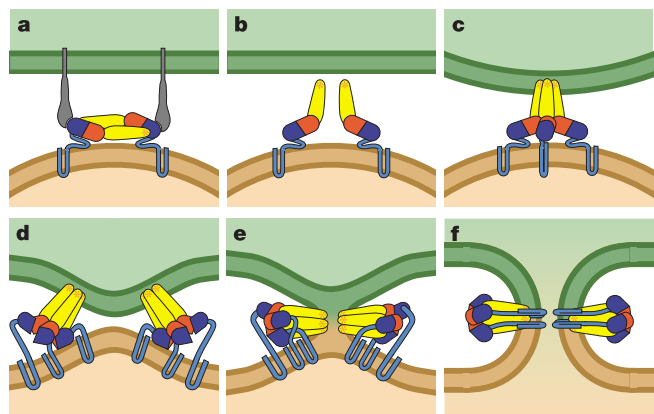


Figure 5 Proposed mechanism for fusion mediated by class II viral fusion proteins. Full-length E is represented as in Fig. 1c, with the stem and viral transmembrane anchor in cyan. **a**, E binds to a receptor on the cell surface and the virion is internalized to an endosome. **b**, Reduced pH in the endosome causes domain II to hinge outward from the virion surface, exposing the fusion loop, and allowing E monomers to rearrange laterally in the plane of the viral membrane. **c**, The fusion loop inserts into the hydrocarbon layer of the host-cell membrane, promoting trimer formation. **d**, Formation of trimer contacts spreads from the fusion loop at the tip of the trimer, to the base of the trimer. Domain III shifts and rotates to create trimer contacts, causing the C-terminal portion of E to fold back towards the fusion loop. Energy release by this refolding bends the apposed membranes. **e**, Creation of additional trimer contacts between the stem-anchor and domain II leads first to hemifusion and then **(f)** to formation of a lipidic fusion pore.

have no particular sequence conservation. Indeed, the Ebola virus fusion peptide begins at the 23rd residue of GP2, rather than at the N terminus³⁹, and a cysteine preceding the fusion peptide probably makes a disulphide bond with a cysteine C terminal to it. Thus, whatever its conformation, this peptide must enter and leave the membrane from the external face. The available data for class I fusion peptides are thus consistent with one important feature of our structure and of the SFV E1 postfusion structure¹⁷: insertion only into the outer bilayer leaflet.

Insertion only into the outer leaflet is also consistent with the requirement of a complete C-terminal transmembrane anchor on influenza HA or Simian virus 5 for full fusion to take place^{40–42}. Indeed, as illustrated by Fig. 5f, one of the two membrane attachment structures must span the bilayer to stabilize a fusion pore. This appears to be the C-terminal anchor for both class I and class II fusion.

Strategies for inhibiting flavivirus fusion

The discovery of a hydrophobic ligand-binding pocket beneath the kl loop in the prefusion structure of dengue sE has suggested one possible strategy for inhibiting flavivirus entry: by interfering with the fusion transition¹⁰. The rationale for that proposal is enhanced by our new structure, which shows that significant rearrangements do occur around the kl loop during the conformational change. The trimer structure also suggests a second strategy for interfering with fusion, related to an approach successful in developing an HIV antiviral compound⁴³. Peptides corresponding to the C-terminal region of the gp41 ectodomain inhibit HIV-1 entry, probably by binding to the trimeric, N-terminal 'inner core' of the protein and interfering with the folding back against it of the C-terminal 'outer layer'⁴⁴. An analogous strategy may be successful with class II viral fusion proteins, such as those of dengue and perhaps of hepatitis C. The way in which the stem is likely to fold back suggests that peptides derived from stem sequences could block completion of the conformational change, by interacting with surfaces on the clustered domains II. This would interfere with the final stage of the conformational change, whereas targeting the pocket beneath the kl loop would interfere with the first stage. Inhibitors developed from these two strategies could thus act in synergy. □

Methods

Expression and purification of dengue sE

Soluble E protein (sE) from dengue virus type 2 S1 strain⁴⁵ was supplied by Hawaii Biotech. The protein was expressed in *Drosophila melanogaster* Schneider 2 cells (obtained from ATCC) using a pMtt vector (GlaxoSmithKline) containing the dengue 2 prM and E genes (nucleotides 539–2121), as previously described⁴⁶. The resulting prM-E preprotein is processed during secretion to yield sE, which was purified from the cell culture medium by immunoaffinity chromatography⁴⁷.

Preparation of dengue sE trimers

Dengue sE trimers were obtained as follows, based on a method developed for TBE virus sE¹⁸. 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (Avanti Polar Lipids) and 1-cholesterol (Sigma) were dissolved in chloroform, mixed in a 1:1:1 molar ratio, and dried under high vacuum for over 4 h. The lipid film was resuspended in 10 mM triethanolamine (TEA) of pH 8.3, 0.14 M NaCl and subjected to five cycles of freeze-thawing, followed by 21 cycles of extrusion through two 0.2 µm polycarbonate filter membranes (Whatman). Purified sE was added in a 1:680 protein:lipid molar ratio and incubated at 37 °C for 5 min. The pH was lowered to endosomal levels by adding 75 mM MES of pH 5.4, and the protein was incubated at 37 °C for 30 min. Liposomes were solubilized with a 20-fold molar excess of *n*-octyl-β-D-glucoside (β-OG) and 4 mM *n*-undecyl-β-D-maltoside (UDM) (Anatrace). Excess lipid was removed by cation exchange chromatography with MonoS (Pharmacia) in 25 mM citric acid of pH 5.26, 70 mM NaCl, 4 mM UDM. After washing with 0.4 M NaCl, the protein was eluted with 1–1.5 M NaCl. E trimers were further purified by gel filtration on a Superdex 200 column (Pharmacia) in 8 mM TEA of pH 7, 80 mM NaCl, 3 mM UDM. The sE trimers were concentrated to about 15 mg ml⁻¹ for crystallization and dialysed against the gel filtration buffer using a 50-kDa molecular-mass cutoff membrane (Spectrapor).

Crystallization and data collection

Crystals were grown at 20 °C by hanging-drop vapour diffusion by mixing equal volumes of protein solution and the following reservoir solution: 20–30% polyethylene glycol 400

(PEG400), 0.1 M MOPS pH 7–8 or Tris pH 8–9, 80 mM NaCl. Two crystal forms were obtained: plates of space group *P*321 with cell dimensions $a = b = 76.2 \text{ \AA}$, $c = 131 \text{ \AA}$, and rhomboids of space group *P*3₂21 with cell dimensions $a = b = 153 \text{ \AA}$, $c = 143 \text{ \AA}$. The asymmetric unit of the *P*321 crystals contains one molecule of sE; that of the *P*3₂21 crystals, one trimer (three molecules) of sE. Cryoprotection was achieved by raising the concentration of PEG400 to 30%. Crystals were frozen in liquid nitrogen, and all data were collected at 100 K on BioCARS beamline 14-BM-C at the Advanced Photon Source (Argonne National Laboratory). The data were processed with HKL⁴⁸. Data collection statistics are presented in Supplementary Table 1.

Chemical cross-linking

sE was covalently cross-linked with ethylene glycol *bis*-(succinimidyl succinate) (EGS). 10 µM–1 mM EGS was added from a fresh 0.1 M stock solution in dimethyl sulphoxide to about 5 µg E at 10 µg ml⁻¹. Acidic solutions were neutralized with TEA of pH 8.5. After 30 min at room temperature, EGS was quenched with 20 mM Tris for 15 min. Protein was precipitated with trichloroacetic acid and resuspended in SDS–polyacrylamide gel electrophoresis (PAGE) sample buffer for gel electrophoresis.

Electron microscopy

Dengue sE trimers inserted into liposomes were prepared as described above and adsorbed to glow-discharged, carbon-coated copper grids. Samples were washed with two drops of deionized water, stained with two drops of 0.7% uranyl formate for 20 s, washed with water, and blotted gently. Micrographs were recorded on a Philips Tecnai 12 electron microscope at 100 kV and 64,000-fold magnification.

Structure determination

The crystal structure of dengue E in the postfusion conformation was determined by molecular replacement using individual domains from the prefusion dengue E structure¹⁰ (Protein Data Bank code 1OKE) as search models, and the *P*321 data set (see Supplementary Table 1). Domain II was placed first, followed by domain I, with AmoRe⁴⁹. Domain III was placed last, with CNS⁵⁰. The atomic coordinates of the three domains were refined as rigid bodies. The model was rebuilt with O⁵¹ based on $2F_o - F_c$ and $F_o - F_c$ Fourier maps. Residues 1–17, 34–40, 49–54, 128–137, 165–192, 290–299 and 341–346 were built *de novo*. Coordinates were then refined against data up to 2.0 Å resolution by simulated annealing using torsion-angle dynamics with CNS⁵⁰, and rebuilt with O, in iterative cycles. Later cycles included restrained refinement of B-factors for individual atoms and energy minimization against maximum likelihood targets with CNS. The final model contains residues 1–144 and 159–394, an *n*-acetyl glucosamine glycan on residue 67, 205 water molecules and one chloride ion. Residues 145–158 and the glycan on residue 153 are disordered. Refinement statistics are presented in Supplementary Table 1.

Received 16 August; accepted 24 October 2003; doi:10.1038/nature02165.

1. Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: the influenza haemagglutinin. *Annu. Rev. Biochem.* **69**, 531–569 (2000).
2. Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373 (1981).
3. Baker, K. A., Dutch, R. E., Lamb, R. A. & Jardetzky, T. S. Structural basis for paramyxovirus-mediated membrane fusion. *Mol. Cell* **3**, 309–319 (1999).
4. Melikyan, G. B. *et al.* Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion. *J. Cell Biol.* **151**, 413–423 (2000).
5. Russell, C. J., Jardetzky, T. S. & Lamb, R. A. Membrane fusion machines of paramyxoviruses: capture of intermediates of fusion. *EMBO J.* **20**, 4024–4034 (2001).
6. Bullock, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
7. Chen, J., Skehel, J. J. & Wiley, D. C. N- and C-terminal residues combine in the fusion-pH influenza haemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc. Natl Acad. Sci. USA* **96**, 8967–8972 (1999).
8. Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C. & Harrison, S. C. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* **375**, 291–298 (1995).
9. Lescar, J. *et al.* The fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* **105**, 137–148 (2001).
10. Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc. Natl Acad. Sci. USA* **100**, 6986–6991 (2003).
11. Allison, S. L. *et al.* Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. *J. Virol.* **69**, 695–700 (1995).
12. Ferlenghi, I. *et al.* Molecular organization of a recombinant subviral particle from tick-borne encephalitis virus. *Mol. Cell* **7**, 593–602 (2001).
13. Kuhn, R. J. *et al.* Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717–725 (2002).
14. Allison, S. L., Schlich, J., Stiasny, K., Mandl, C. W. & Heinz, F. X. Mutational evidence for an internal fusion peptide in flavivirus envelope protein E. *J. Virol.* **75**, 4268–4275 (2001).
15. Levy-Mintz, P. & Kielian, M. Mutagenesis of the putative fusion domain of the Semliki Forest virus spike protein. *J. Virol.* **65**, 4292–4300 (1991).
16. Ahn, A., Gibbons, D. L. & Kielian, M. The fusion peptide of Semliki Forest virus associates with sterol-rich membrane domains. *J. Virol.* **76**, 3267–3275 (2002).
17. Gibbons, D. L. *et al.* Conformational change and protein–protein interactions of the fusion protein of Semliki Forest virus. *Nature* **427**, 320–325 (2004).
18. Stiasny, K., Allison, S. L., Schlich, J. & Heinz, F. X. Membrane interactions of the tick-borne encephalitis virus fusion protein E at low pH. *J. Virol.* **76**, 3784–3790 (2002).
19. Wimley, W. C. & White, S. H. Partitioning of tryptophan side-chain analogs between water and cyclohexane. *Biochemistry* **31**, 12813–12818 (1992).
20. Allison, S. L., Stiasny, K., Stadler, K., Mandl, C. W. & Heinz, F. X. Mapping of functional elements in the stem-anchor region of tick-borne encephalitis virus envelope protein E. *J. Virol.* **73**, 5605–5612 (1999).

21. Zhang, W. *et al.* Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. *Nature Struct. Biol.* **10**, 907–912 (2003).
22. Crill, W. D. & Roehrig, J. T. Monoclonal antibodies that bind to domain III of dengue virus E glycoprotein are the most efficient blockers of virus adsorption to Vero cells. *J. Virol.* **75**, 7769–7773 (2001).
23. Jennings, A. D. *et al.* Analysis of a yellow fever virus isolated from a fatal case of vaccine-associated human encephalitis. *J. Infect. Dis.* **169**, 512–518 (1994).
24. Lobigs, M. *et al.* Host cell selection of Murray Valley encephalitis virus variants altered at an RGD sequence in the envelope protein and in mouse virulence. *Virology* **176**, 587–595 (1990).
25. Holzmann, H., Heinz, F. X., Mandl, C. W., Guirakhoo, F. & Kunz, C. A single amino acid substitution in envelope protein E of tick-borne encephalitis virus leads to attenuation in the mouse model. *J. Virol.* **64**, 5156–5159 (1990).
26. Jiang, W. R., Lowe, A., Higgs, S., Reid, H. & Gould, E. A. Single amino acid codon changes detected in louping ill virus antibody-resistant mutants with reduced neurovirulence. *J. Gen. Virol.* **74**, 931–935 (1993).
27. Gao, G. F., Hussain, M. H., Reid, H. W. & Gould, E. A. Identification of naturally occurring monoclonal antibody escape variants of louping ill virus. *J. Gen. Virol.* **75**, 609–614 (1994).
28. Cecilia, D. & Gould, E. A. Nucleotide changes responsible for loss of neuroinvasiveness in Japanese encephalitis virus neutralization-resistant mutants. *Virology* **181**, 70–77 (1991).
29. Chen, Y. *et al.* Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nature Med.* **3**, 866–871 (1997).
30. Navarro-Sanchez, E. *et al.* Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses. *EMBO Rep.* **4**, 1–6 (2003).
31. Tassaneetrithep, B. *et al.* DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *J. Exp. Med.* **197**, 823–829 (2003).
32. Stiasny, K., Allison, S. L., Marchler-Bauer, A., Kunz, C. & Heinz, F. X. Structural requirements for low-pH-induced rearrangements in the envelope glycoprotein of tick-borne encephalitis virus. *J. Virol.* **70**, 8142–8147 (1996).
33. Chan, D. C. & Kim, P. S. HIV entry and its inhibition. *Cell* **93**, 681–684 (1998).
34. Kuzmin, P. I., Zimmerberg, J., Chizmadzhev, Y. A. & Cohen, F. S. A quantitative model for membrane fusion based on low-energy intermediates. *Proc. Natl Acad. Sci. USA* **98**, 7235–7240 (2001).
35. Razinkov, V. I., Melikyan, G. B. & Cohen, F. S. Hemifusion between cells expressing hemagglutinin of influenza virus and planar membranes can precede the formation of fusion pores that subsequently fully enlarge. *Biophys. J.* **77**, 3144–3151 (1999).
36. Kozlov, M. M. & Chernomordik, L. V. A mechanism of protein-mediated fusion: coupling between refolding of the influenza hemagglutinin and lipid rearrangements. *Biophys. J.* **75**, 1384–1396 (1998).
37. Wahlberg, J. M., Bron, R., Wilschut, J. & Garoff, H. Membrane fusion of Semliki Forest virus involves homotrimers of the fusion protein. *J. Virol.* **66**, 7309–7318 (1992).
38. Han, X., Bushweller, J. H., Cafiso, D. S. & Tamm, L. K. Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin. *Nature Struct. Biol.* **8**, 715–720 (2001).
39. Ito, H., Watanabe, S., Sanchez, A., Whitt, M. A. & Kawakita, Y. Mutational analysis of the putative fusion domain of Ebola virus glycoprotein. *J. Virol.* **73**, 8907–8912 (1999).
40. Kemble, G. W., Danieli, T. & White, J. M. Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. *Cell* **76**, 383–391 (1994).
41. Armstrong, R. T., Kushnir, A. S. & White, J. M. The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition. *J. Cell Biol.* **151**, 425–437 (2000).
42. Dutch, R. E. & Lamb, R. A. Deletion of the cytoplasmic tail of the fusion protein of the paramyxovirus simian virus 5 affects fusion pore enlargement. *J. Virol.* **75**, 5363–5369 (2001).
43. Baldwin, C. E., Sanders, R. W. & Berkhout, B. Inhibiting HIV-1 entry with fusion inhibitors. *Curr. Med. Chem.* **10**, 1633–1642 (2003).
44. Kilby, J. M. *et al.* Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nature Med.* **4**, 1302–1307 (1998).
45. Hahn, Y. S. *et al.* Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. *Virology* **162**, 167–180 (1988).
46. Ivy, J., Nakano, E. & Clements, D. Subunit immunogenic composition against dengue infection. US Patent 6,165,477 (1997).
47. Cuzzubbo, A. J. *et al.* Use of recombinant envelope proteins for serological diagnosis of dengue virus infection in an immunochromatographic assay. *Clin. Diagn. Lab. Immunol.* **8**, 1150–1155 (2001).
48. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
49. Navaza, J. Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D* **57**, 1367–1372 (2001).
50. Brünger, A. T. *et al.* Crystallography NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
51. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank staff at BioCARS beamline 14-BM-C at the Advanced Photon Source (Argonne National Laboratory). We thank J. Zimmerberg, F. Rey and F. Heinz for discussions, and T. Walz and Y. Cheng for guidance on EM experiments. This work was supported by a long-term fellowship to Y.M. from the Human Frontier Science Program Organization, and by an NIH grant to S.C.H., who is a Howard Hughes Medical Institute Investigator.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.C.H. (harrison@crystal.harvard.edu). Coordinates have been deposited in the Protein Data Bank under accession code 1OK8.

Conformational change and protein–protein interactions of the fusion protein of Semliki Forest virus

Don L. Gibbons^{1,2}, Marie-Christine Vaney¹, Alain Roussel^{1*}, Armelle Vigouroux¹, Brigid Reilly², Jean Lepault¹, Margaret Kielian² & Félix A. Rey¹

¹Virologie Moléculaire & Structurale, UMR 2472/1157 CNRS-INRA, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

²Department of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA

* Present address: AFMB, CNRS UMR 6098, 31 Chemin Joseph Aiguier, 13402 Marseilles, France

Fusion of biological membranes is mediated by specific lipid-interacting proteins that induce the formation and expansion of an initial fusion pore. Here we report the crystal structure of the ectodomain of the Semliki Forest virus fusion glycoprotein E1 in its low-pH-induced trimeric form. E1 adopts a folded-back conformation that, in the final post-fusion form of the full-length protein, would bring the fusion peptide loop and the transmembrane anchor to the same end of a stable protein rod. The observed conformation of the fusion peptide loop is compatible with interactions only with the outer leaflet of the lipid bilayer. Crystal contacts between fusion peptide loops of adjacent E1 trimers, together with electron microscopy observations, suggest that in an early step of membrane fusion, an intermediate assembly of five trimers creates two opposing nipple-like deformations in the viral and target membranes, leading to formation of the fusion pore.

Membrane rearrangements take place throughout the lifespan of a cell and are essential to all forms of life¹. Stable membranes do not fuse spontaneously, and specific fusion proteins tightly control all membrane fusion events through interactions with lipids and with other fusion proteins. Much of our knowledge of membrane fusion has come from the study of viral membrane proteins that interact with lipids to induce fusion during viral entry. The best-studied example of a membrane fusion protein is the influenza virus haemagglutinin, a prototypic class I fusion protein². Although many features of the molecular mechanisms of membrane fusion are not understood, the class I studies suggest a model in which fusion is driven by the formation of a stable protein rod anchored in both the target and donor membranes³. Several such hairpin-like structures could act cooperatively^{4–6} to induce the formation of a fusion pore connecting the two membrane-bound compartments^{7,8}, although the mechanisms by which such synergistic interactions might drive fusion are not clear.

Semliki Forest virus (SFV), a member of the alphavirus genus, has been used as a model system to define virus endocytic entry⁹. Alphaviruses have two type I glycoproteins anchored at their surface, E1 (membrane fusion protein) and E2 (receptor binding protein), organized with a $T = 4$ icosahedral architecture¹⁰. The crystal structure of the neutral-pH form of the E1 ectodomain has been determined¹¹ and fitted into medium-resolution (9 Å) cryo-electron microscopy (EM) reconstructions of SFV¹¹ and Sindbis virus particles¹². These studies have shown that E1 lies in a tangential orientation at the viral surface, forming an icosahedral scaffold covering the viral membrane, whereas E2 forms projecting spikes at the three-fold and quasi-three-fold axes of the $T = 4$ icosahedral lattice of the viral surface. In this arrangement, E2 prevents E1–E1 contacts about three-fold axes of the particle. The crystal structure of E1 revealed an unanticipated homology with the envelope protein E of flaviviruses^{13,14}, with identical topology and a very similar three-domain arrangement of the polypeptide chain. These ‘class II’ membrane fusion proteins¹¹ are folded mostly as β -sheets, with the internal fusion peptide (referred to here as the ‘fusion loop’ or ‘cd loop’) in a loop between two β -strands. During virus entry, the acidic environment of the endosome triggers an

irreversible conformational change in E1, from an E1/E2 heterodimer at the viral surface to a target-membrane-inserted E1 homotrimer¹⁵. This rearrangement of E1 leads to fusion of the viral and endosomal membranes, releasing the virus genomic RNA into the cytoplasm of the cell¹⁶.

The E1 ectodomain has been shown to behave similarly to its full-length, virus-anchored counterpart^{17,18}. This fragment, termed E1* and lacking amino acids 392–438 at the carboxy terminus of the protein, changes from a soluble monomer to a trimer as it inserts into liposomes upon exposure to low pH¹⁷. E1* membrane insertion exhibits the same lipid dependence as alphavirus fusion, requiring both cholesterol and sphingolipid in the target membrane¹⁹. Electron micrographs of liposomes containing E1 or E1* inserted through low-pH treatment showed that both proteins reorient vertically with respect to the membrane to produce trimers of very similar shape and dimensions²⁰. Although the depth of membrane insertion was not certain, the images suggested a possible lipid-bilayer-spanning topology for domain II (see Fig. 1 for the definition of protein domains), with the fusion loop and the transmembrane (TM) anchor at the same side of a stable protein rod, similar to the hairpin structure of post-fusion class I proteins.

Here, we have solubilized the E1* trimer from liposomes, crystallized it in the presence of detergents and determined its three-dimensional (3D) structure to 3.3 Å resolution. This structure shows that the low-pH-triggered trimeric form of E1* has the C-terminal ‘stem’ region, which connects to the TM anchor, directed towards the fusion loop. This observation therefore confirms that a folded-back conformation is a common fusion mechanism for both class I and class II proteins. However, in contrast to the class I proteins²¹, the three domains of E1 maintain their original folds during this conformational change. The E1* fusion loop has an extended conformation, consistent with insertion into only the outer leaflet of the target membrane. The contacts observed in the crystals, along with the symmetry of the rosette-like structures observed by EM, suggest that fusion involves an assembly of several E1 trimers interacting through their fusion loops around the fusion site. These lateral interactions between fusion peptide loops from

adjacent trimers are postulated to have an essential and concerted role in the fusion process.

Tertiary rearrangement of the E1 ectodomain

Figure 1 shows the observed conformational rearrangement of the E1* polypeptide chain induced by low-pH treatment in the presence

of liposomes. There is essentially no change in secondary structure. By contrast, the interactions between the domains constituting the E1 subunit are markedly different. The most important rearrangement concerns domain III, which moves by 37 Å towards the fusion loop. The different location and orientation of domain III redirects the polypeptide chain so that in the trimeric form its C terminus

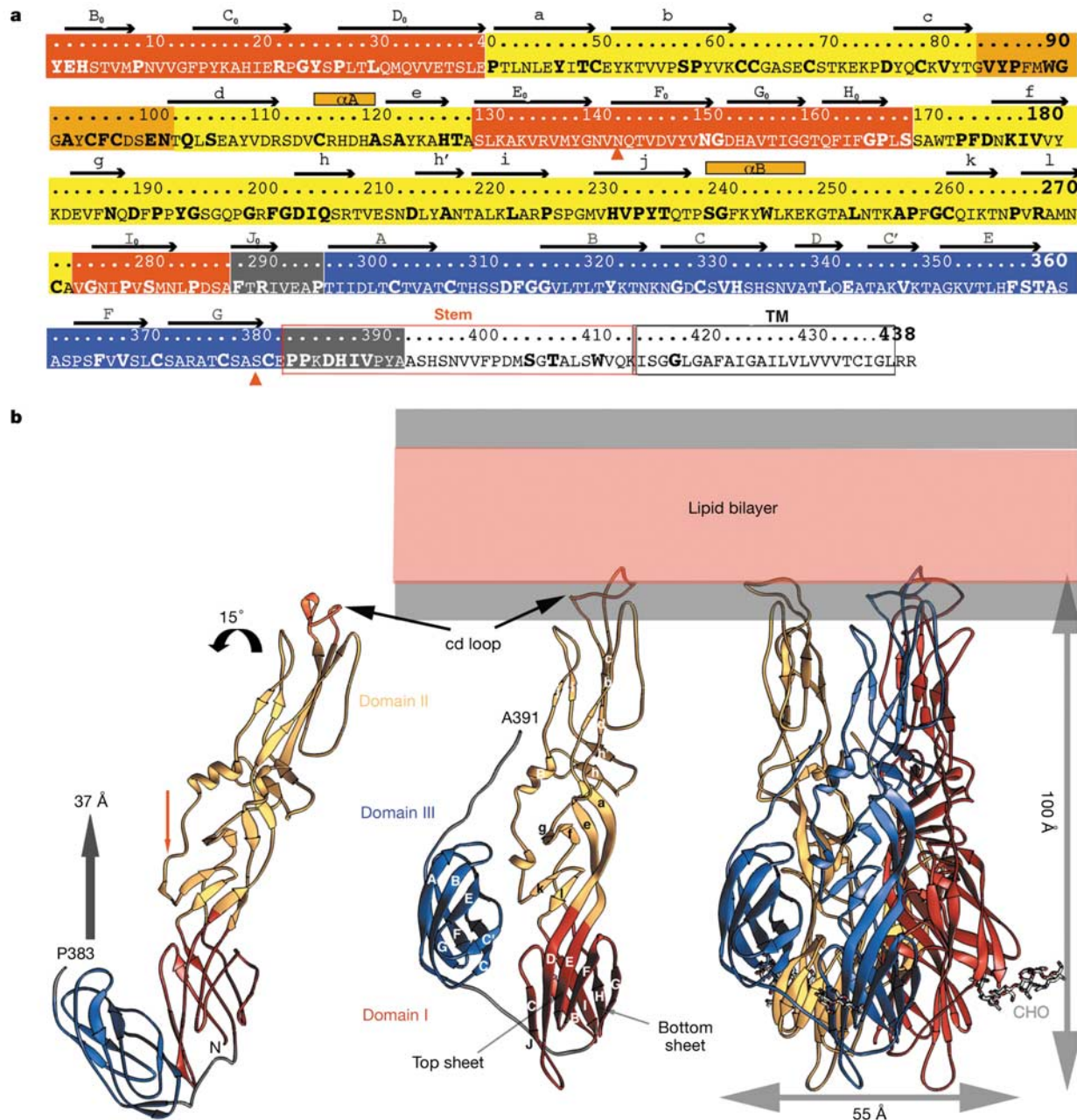


Figure 1 Overall fold of glycoprotein E1. **a**, Amino-acid sequence of SFV E1. Highly conserved residues are highlighted as bold letters in bigger font. Red, yellow and blue backgrounds denote residues belonging to domains I, II and III, respectively. The fusion loop is against an orange background, and the stem and the linker between domains I and III are in grey. The stem and transmembrane anchor are boxed in red and black, respectively. A white background denotes residues absent from the crystallized fragment. Elements of secondary structure (labelled as in ref. 13) are indicated above the sequence. Red triangles underneath the sequence indicate glycosylation sites: N-linked at position 141 and O-linked at position 379. **b**, Left panel, ribbon diagram of the E1 ectodomain at neutral pH. Thick arrows indicate the major rearrangements. A thin red arrow points to a region of contact described in the text. The N and C termini of the fragment are indicated

by N and P383 (last ordered residue), respectively. Middle panel, the E1* subunit after the conformational change, with elements of the secondary structure labelled. The two forms (left and middle panel) were superposed on the B₀L₀H₀G₀ β-sheet (labelled 'bottom sheet') in domain I. A391 indicates the C terminus of the fragment. A new β-strand, J₀, runs antiparallel to β-strand C₀ in domain I. A sketch of the lipid bilayer, at roughly the same scale as the protein, is indicated at the top, where grey and pink regions indicate lipid heads and lipid aliphatic chain regions, respectively. Right panel, the E1* trimer. Each subunit is coloured differently, with the subunit in blue oriented as in the middle panel. The molecular dimensions are indicated. The N-linked carbohydrate is displayed as grey balls and sticks, labelled CHO.

points towards the fusion loop region. A segment of about nine amino acids of the stem (at the extreme C terminus of the E1* fragment), which was not ordered in the crystals of the monomeric neutral-pH form, becomes ordered in the trimer. The E1* fragment ends at residue 391, which lies about 25 Å from the fusion peptide loop. The remaining residues connecting to the TM anchor, 392–413, can easily cover the distance to the lipid bilayer, indicated by the location of the fusion loop. In the trimer, the immunoglobulin-like domain III interacts through its AB loop and strand C' (see Fig. 1 for nomenclature) with amino acids 250–260 of domain II, which lie in a region of relatively extended conformation in the monomeric neutral-pH form (indicated by a red arrow in Fig. 1b, left panel). The contact induces a slight winding of this domain II segment, which packs against strand g and the kl hairpin. A consequence of this interdomain interaction is to pull down helix α B, bringing with it the tip of domain II, which rotates by 15° about a hinge located between β -sheets kD₀E₀ and gfeah (indicated in Fig. 1b), resulting in a straight continuous rod containing domains I and II. This interaction is stabilized by the fragment of the stem present in the structure (grey in Fig. 1).

The fusion cd loop (orange in Fig. 1) displays a higher average temperature factor than the rest of the molecule. It unwinds with respect to the neutral-pH soluble form, displaying a rather extended conformation that is not compatible with penetration of the aliphatic moiety of the membrane. It exposes several aromatic side chains that are very probably interacting with disordered

detergent molecules in the crystal. Assuming that the observed conformation of the fusion loop does indeed correspond to its membrane-interacting conformation, our data suggest that the glycine-rich main chain interacts tightly with the lipid heads, projecting aromatic side chains into the aliphatic region of the lipid bilayer. This loop shows considerable plasticity, adapting to the particular contact in the crystal, as discussed below.

Overall architecture of the E1* trimer

The subunits associate in a parallel arrangement in the trimer, such that the fusion loops are at the same side of a stable elongated molecule, as shown in Fig. 1b. Each subunit in the E1* trimer buries an area of 4,300 Å² from solvent accessibility. The extent of the contacts is reflected in the observed stability of the molecule^{18,22}. The trimer is formed essentially by the central interactions of the bottom β -sheet of domain I (formed by strands G₀H₀I₀B₀, see Fig. 1) and is continued in domain II by contacts between elements at its domain I proximal half. The latter involve, on one side, residues spanning helix α B all the way to the beginning of β -strand k and, on the other, β -sheet gfeah. Domain III packs at the periphery, at the interface between two subunits. The highly conserved residues in alpha-viruses cluster at the fusion loop, at the stem and at the trimer contact regions. Both the trimer surface and the intra-trimer contact areas are essentially hydrophilic. The tips of domain II do not display trimer contacts, and exhibit high temperature factors. The location of the C terminus of the ectodomain suggests that in the final, post-fusion conformation of the full-length trimer, the remaining amino acids of the stem are likely to provide a contact surface to stabilize this region.

Lateral interactions of E1 trimers on membranes

Before the determination of the crystal structure, we derived a model for the E1* trimer from EM data²⁰. This model was based on

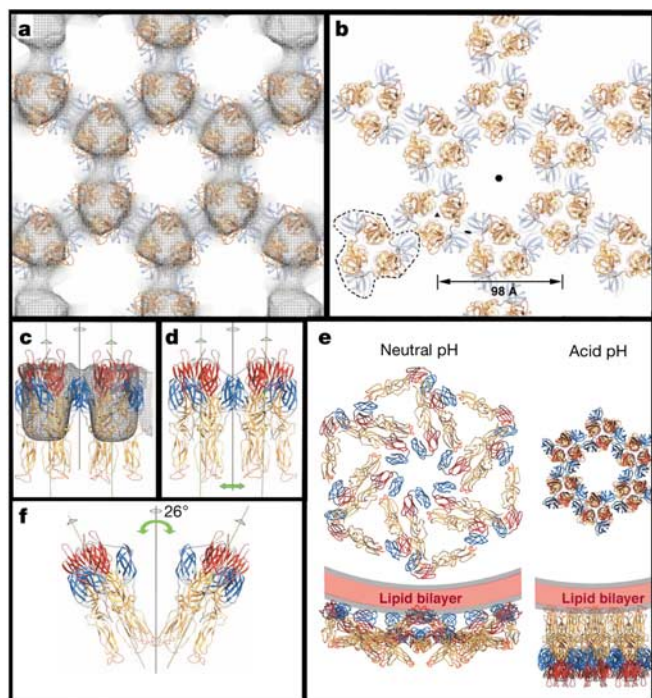


Figure 2 Trimer-trimer interactions observed in the 2D lattice and in the 3D crystals. **a–d**, E1* trimers on the 2D lattice²⁰. **a** and **b** show views from the membrane, **c** and **d** show side views. In **b** and **d**, the reconstruction was removed to visualize the contacts. In **b**, a trimer is outlined with a black broken line. The $p6$ hexagonal cell parameter is indicated (98 Å) and one of each symmetry element (six-fold, three-fold and two-fold axis) is marked with a filled black symbol: hexagon, triangle and ellipse, respectively. In **c** and **d**, the symmetry axes are shown as vertical lines with open symbols. In **d**, a green double arrow indicates that the fusion loops are too far apart to make contacts. **e**, Conformational rearrangement of the virus surface. Left panel, arrangement of E1 at the viral surface at neutral pH^{11,12} (glycoprotein E2 is not shown). Right panel, arrangement of E1* after the conformational change, as observed at the surface of liposomes. First row, top view; second row, side view. **f**, Two-fold contact on the 3D crystal, involving fusion loops. The angle between trimers and dyad is 26°.

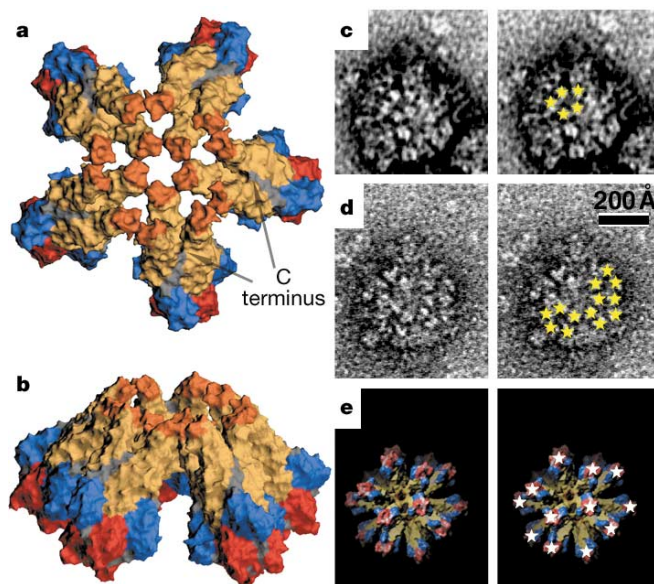


Figure 3 A ring of five trimers of E1*. **a**, **b**, Surface representation of the ring, coloured according to domains, view down the five-fold axis (**a**) and a side view (**b**) slightly tilted to show the depth of the crater. Note the grey stem segment leading to the C terminus of the fragment, indicated by arrows. **c**, **d**, Micrographs of negatively stained rosettes of E1* trimers obtained by dialysing away the detergent used for solubilization. In the right-hand panel, a star was added to highlight the trimers present in rings of five. The three-fold symmetry of some of the trimers is evident, especially on panel **c**. **e**, Model for a symmetric rosette exhibiting dodecahedral symmetry, built according to Supplementary Information. The magnification is about three times that of the rosettes presented in panels **c** and **d**.

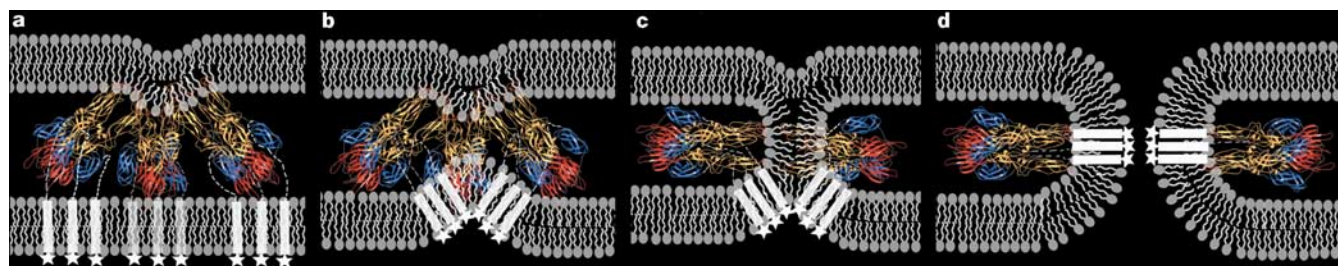


Figure 4 Model for membrane fusion involving protein–protein interactions, as explained in the text. **a–d**, White stars denote hydrophilic residues and C-terminal charge in the cytosolic tail of the fusion protein. White cylinders denote transmembrane anchors.

Broken white lines indicate the stem region, which connects the ectodomain to the TM segments. For clarity, the complete ring of five trimers is not shown.

fitting three E1* subunits, in their neutral-pH form, into a 3D reconstruction calculated from images of negatively stained two-dimensional (2D) arrays of E1*. Because of the low resolution of that reconstruction, we made no attempt at the time to alter the relative orientations of domains within the subunits. The resulting trimer model predicted the observed parallel arrangement of the subunits, but was longer than the actual structure of the E1* trimer because the rearrangement of domain III was not taken into account. Indeed, on the basis of cryo-EM observations of E1*-containing liposomes, we had speculated that domain II might actually cross the lipid bilayer. The predicted length of the trimer was roughly consistent with the height of the extramembrane domain plus a membrane-inserted domain of approximately the thickness of the lipid bilayer. Here we revise the model for E1* membrane insertion by taking into account the fact that the observed conformation of domain II in the crystal structure is incompatible with a membrane-spanning topology, and that the total length of the molecule is shorter because domain III is displaced towards the fusion loop. To better understand the interactions of E1* trimers in the 2D arrays, we have now fitted the experimentally determined structure of the E1* trimer into the previous 3D reconstruction, as described in Methods. As shown in Fig. 2, the trimers make lateral contacts through dyads of the *p6* lattice. The side view (Fig. 2c) shows that the EM reconstruction imaged only the thicker portion of the molecule. The three legs of the ‘tripod’ (that is, the tip of domain II) were not visualized, presumably owing to the reduced mass and increased flexibility caused by the absence of subunit contacts within this portion of the trimer. The trimers display a parallel arrangement in the 2D lattice, with the heads touching but with the ‘legs’ too far apart to make inter-trimer contacts. However, the extensive contacts between fusion loops from adjacent trimers observed in both the 3D crystals and the trimer rosettes (as explained in the following paragraphs) suggest that the fusion loop interaction is likely to be maintained in the hexagonal 2D lattice. Such fusion loop contacts would force the

playing of the domain II legs, allowing the trimer to make contacts through both the head and fusion peptide loops in the 2D lattice. It is noteworthy that the full-length E1 trimer does not form regular lattices on membranes, suggesting that the presence of the stem region stabilizes the legs and prevents the parallel trimer arrangement in a hexagonal lattice (see Supplementary Information).

Figure 2e compares the interactions of E1* trimers in the hexagonal lattice with the interactions of the neutral-pH E1 on the virus surface. In both panels (neutral and acid pH) the same number of E1* subunits is displayed (18 subunits). The surface covered by E1 in the neutral-pH form is more than three times larger than that covered by its low-pH-induced trimeric form. During infection, it is likely that only the E1 molecules most proximal to the endosome membrane can actually interact with it. Rearrangement of the remaining E1 molecules in the virus particle may have a further role in destabilizing the virion membrane to facilitate the membrane fusion process.

Inter-trimer contacts through fusion loops

Insertion of E1 into target membranes seems to be a cooperative process, and depends critically on the lipid composition of the target membrane^{19,20}. Furthermore, the membrane fusion reaction triggered by viral fusion proteins in general is thought to involve interactions between several trimers^{4,5}, which may interact through the fusion loops at the same time as they interact with membranes. In this first structure of a membrane fusion protein containing the membrane-interacting form of the fusion peptide, it is important to examine the possible types of contact between trimers that are mediated by the fusion loop. Figure 2f shows the main contact between trimers in the crystal, termed contact 1 in Supplementary Information. This contact is made across a two-fold axis, but with the trimer axis tilted 26° with respect to the dyad, instead of being parallel as in the 2D lattice. The same surface of the trimer faces the dyad in both the 3D and the 2D crystals, as deduced from the fitting to the EM density. Contact 1 takes place exclusively through the

Table 1 Phasing statistics as a function of resolution*

Resolution range	Completeness (%)	Signal-to-noise ratio	Isomorphous phasing power (centric reflections)†	Isomorphous phasing power (acentric reflections)†	Anomalous phasing power‡	Overall figure of merit (acentric)‡	Overall figure of merit (centric)‡
50.14–8.91	99.9	28.03	0.99	1.34	2.28	0.85	0.55
8.91–6.35	99.9	25.00	1.08	1.56	2.90	0.82	0.52
6.35–5.20	99.8	23.91	1.09	1.48	2.82	0.76	0.48
5.20–4.51	99.4	23.45	0.77	1.14	2.14	0.72	0.44
4.51–4.04	99.4	17.81	0.73	0.92	1.61	0.62	0.41
4.04–3.69	99.0	11.97	0.60	0.64	0.99	0.45	0.26
3.69–3.42	98.9	10.56	0.45	0.48	0.77	0.34	0.18
3.42–3.20	71.9	8.49	0.57	0.49	0.83	0.22	0.15
Overall	96.0	15.17	0.91	1.07	1.81	0.54	0.38

Refinement statistics: resolution range, 20–3.2 Å; number of reflections, 40,912; working set, 38,837 reflections; *R* factor, 0.265; free set, 2,075 reflections; free *R* factor, 0.285. r.m.s.d. bonds, 0.010006 Å; r.m.s.d. angles, 1.6°. r.m.s.d. domain superposition of neutral/acid-pH forms: domain I, 2.43 Å; domain II, 1.76 Å; domain III, 1.0 Å.

*Statistics calculated using the program SHARP²⁷.

†The phasing power statistics corresponds to each crystal separately. We took a representative crystal to fill these columns.

‡The reported figure of merit corresponds to the phases before any density modification.

glycine-rich, flexible fusion peptide loops. These interactions are therefore likely to adapt to different geometries and, in particular, to different angles between trimers.

If, in contrast to the 3D crystals, where the packing environment of the trimers is not three-fold symmetric, three-fold symmetry is imposed on the two trimers involved in contact 1, it is possible to create a closed arrangement of trimers, the symmetry of which depends on the angle between the trimer axis and the dyad as explained in Supplementary Information. A ring of five trimers making identical interactions best matches the experimentally observed packing angle in the crystal. Figure 3a, b shows such a ring, generated by changing the angle observed in the crystal from 26° to 21°, which is the angle between triads and dyads surrounding the five-fold axes of a dodecahedron. EM observations of E1* rosettes, obtained when the detergent is removed from purified, detergent-solubilized E1* trimers, did indeed reveal the presence of rings of five trimers (Fig. 3c–d). The rosettes are irregular, in keeping with the alternative possible arrangements of three, four or six trimers described in Supplementary Information, but they show the presence of several five-trimer rings. For comparison, Fig. 3e shows a rosette generated from the crystal structure by introducing three identical fusion loop contacts into each of the 20 trimers present in a dodecahedron, with the fusion loops buried in the interior. During the interactions of the trimer with membranes, however, the geometry of the lipid bilayer is likely to restrict the assembly and allow only the formation of one pentagonal face of the dodecahedron, as the other faces would be out of the plane of the membrane.

An intermediate assembly in fusion

The interactions observed between fusion loops, together with the five-trimer rings visualized by EM, suggest that the E1* trimer may correspond to the three-fold symmetric portion of an intermediate made by the full-length E1 molecule while anchored simultaneously to the target and viral membranes. In such an intermediate, the remaining C-terminal stem region, absent in E1*, is likely to be an adaptor in a transient break of three-fold symmetry occurring during the fusion process, whereas the final fully symmetric molecule would have both the stem and TM domains in place. In such a model, the role of this stem segment would be to bring the viral membrane anchor into place in the final fully symmetric molecule. Following insertion of the fusion peptide into the target membrane, this final conformation can only be reached by the trimer within a fused membrane. These E1 conformations together suggest a model for the concerted actions of E1 during fusion, as illustrated in Fig. 4. The ring of five trimers displayed in Fig. 3a, b has a volcano appearance, with a crater formed by five fusion loops at the base and ten at the rim. If each fusion loop is postulated to interact with the lipid heads of the membrane, both at the rim and at the base of the crater, a nipple-like deformation of the target lipid bilayer will be induced, as indicated in Fig. 4a. In this first intermediate, the TM anchors are attached to the viral membrane and the stem has not yet found its expected location in the final post-fusion form of E1. This arrangement will pull the TM segments towards the fusion loop, introducing a second nipple in the viral membrane, exactly under the one formed in the target membrane (Fig. 4b). The stem is expected to interact tightly with the body of the E1* trimer in the final, post-fusion form of the full-length E1 molecule. In this process, the stem will also induce the opposite curvature on the viral membrane while forcing it against the nipple formed in the target membrane, effectively forcing the outer leaflets of the two membranes to merge and form a hemifusion stalk (indicated in Fig. 4c). The presence of hydrophilic residues at the cytoplasmic side, or even just the negative charge of the C terminus of the polypeptide chain (indicated by stars) would prevent the TM segments from entering the stalk. Completion of the conformational change of the protein to reach its final, lowest-energy form, in

which the trimer is fully three-fold symmetric, can therefore be achieved only by forcing formation of a pore (Fig. 4d). The proposed initial aqueous pore would thus be lined by the cytosolic C-terminal end of the fusion protein. This model is in agreement with observations on the human immunodeficiency virus, in which the formation of the six-helix bundle of gp41 was completed only immediately after pore formation²³. It also explains why substitution of the TM anchor in influenza haemagglutinin by a glycosyl phosphatidylinositol (GPI) anchor²⁴, or truncation of the anchor so that it can no longer cross the viral membrane²⁵, leads only to hemifusion. Similar results were reported for C-terminal truncations of the paramyxovirus SV5 fusion protein²⁶. The proposed model may also help to explain in part the observed effect of the cytoplasmic tail of several retroviruses in controlling fusion^{27,28}. The intermediate depicted in Fig. 4c very likely corresponds to the 'restricted hemifusion' state observed in the case of influenza²⁹ and rabies virus³⁰. In such a state, in spite of continuity between the two outer leaflets, the lipids cannot freely diffuse past the ring formed by the TM segments of the fusion protein, resulting in the observed effect. Transition from the step shown in Fig. 4c to that shown in Fig. 4d is likely to involve overcoming a high energy barrier. It has indeed been observed in the case of influenza haemagglutinin that reaching the fusion pore state from the restricted hemifusion state is strongly temperature dependent³¹.

Concluding remarks

Our data, obtained using the independent approaches of X-ray crystallography and EM, provide the first structural evidence for protein–protein interactions between fusion proteins, postulated as essential for membrane fusion. In addition, the results presented in this and in the accompanying manuscript on the dengue virus glycoprotein³², as well as the structure of the tick-borne encephalitis flavivirus envelope protein in its low-pH-induced trimeric form³³, provide a clear link between the class I and class II viral fusion proteins. The analogies in the folded-back conformations and the ability of the SFV membrane fusion model to explain the intermediates observed with class I fusion proteins strongly suggest that both classes act by a single, universal mechanism to cause membrane fusion. □

Methods

Structure determination

The detailed protein preparation and crystallization procedures are described separately³⁴. Briefly, E1*-containing liposomes were prepared as in ref. 20 and were solubilized with *n*-octyl glucoside. The lipids were removed by extensive washing through a vivaspin microfilter, 100 kDa cutoff. For crystallization trials, the detergent was exchanged while concentrating using the same vivaspin microconcentrators. The best crystals grew at 4 °C in the presence of 15 mM of the detergent DDAO (*N,N*-dimethyl decylamine-*N*-oxide), 250 mM NaBr, 25 mM Tris-HCl and 2.5–12.5% PEG 400, pH 4. They belong to the trigonal space group *P*3₁21 with cell parameters *a* = *b* = 198 Å and *c* = 116 Å. For data collection, the crystals were transferred to a cryoprotectant solution containing 30% PEG 400 and 25% glycerol, mounted into cryo-loops and flash-cooled by plunging into liquid ethane. Native crystals diffracted to about 3.7 Å, but heavy-atom soaks with 10 mM HoCl₃ yielded diffraction to 3.1 Å. The best diffracting crystals were subsequently grown in the presence of 10 mM HoCl₃. The structure was solved by a combination of MAD/isomorphous replacement methods using Ho³⁺ as heavy-atom derivative and anomalous scatterer.

The crystals diffracted to about 3.1 Å along *c** and to 3.3 Å along *a**, but were very radiation sensitive and decayed so that the diffraction pattern reached only about 3.9 Å after five to ten images at 100 °K. Diffraction data from 15 crystals (all grown in the presence of Ho³⁺) were therefore combined to obtain a complete and redundant data set (at three wavelengths around the Ho³⁺ edge) to 3.3 Å resolution. Native data to 4 Å from a single crystal were combined with the Ho³⁺ data set to compute isomorphous differences for phasing. The diffraction data were processed using the programs Denzo and Scalepack³⁵. SOLVE³⁶ was used to scale (using the local scale option) data from different crystals and to determine the positions of heavy atoms. The program SHARP³⁷ was used to refine them and to calculate experimental phases to the limit of the diffraction power of the Ho³⁺-containing crystals, approaching 3.1 Å (but the overall signal-to-noise ratio of the data dropped strongly after 3.3 Å resolution, owing to the anisotropic diffraction). The program DM³⁸ was used for density modification with three-fold non-crystallographic averaging. The experimental maps were very clear except for regions of the cd loop, which have high temperature factors. Model building was done using O³⁹ and Turbo⁴⁰.

Refinement was done with CNS⁴¹ using the MLHL target function, with restraints for non-crystallographic symmetry (except for the fusion loop region, where the experimental density showed clear deviation from the three-fold molecular symmetry).

In the final model, the cd loop shows considerable plasticity, adapts to the specific contact in the crystal and displays high temperature factors, reflecting high mobility and/or static disorder. The packing in the crystal explains the very high radiation sensitivity of the crystals. The crystal packing involves four different contacts, two of which, contacts 1 and 2, are made by the fusion loops and are discussed in Supplementary Information. Contact 1 is mediated by a crystallographic dyad and involves fusion loop residues 85–92 of one trimer and 93–96 in the adjacent trimer. Contact 2 relates residues in the bc loop (63–69) to residues 89–92 in the cd loop of the neighbouring trimer. Contact 3 takes place about another crystallographic dyad, relating two short β -strands, which form a short antiparallel β -sheet, between main-chain residues 139–143 at the amino terminus of strand F₀, including an *N*-linked glycosylation site at position 141. The carbohydrates attached to N141 also participate in this contact, resulting in interpretable electron density corresponding to seven sugar residues, including 2 *N*-acetyl glucosamines, a fucose and four mannose residues. Contact 4 involves the C₀D₀ loop (which is at the opposite end of the molecule from the cd loop; see Fig. 2) interacting with the FG loop of domain III. The phasing and crystallographic refinement statistics are provided in Table 1.

Fitting the atomic model to the EM reconstruction

The parameters and symmetry of the 2D hexagonal lattice ($a = 98 \text{ \AA}$, space group $p6$, implying six-, three- and two-fold axes normal to the plane of the lattice) constrain the fit such that the only degrees of freedom are a rotation about the three-fold molecular axis (which coincides with three-fold lattice axis) and a translation along it. As the trimers lie on a plane, this translation does not affect the lateral contacts between them. The pronounced triangular, propeller-like outline of the molecule results in a unique orientation about the three-fold axes in which the trimers can interact without clashing with each other. Figure 3 shows that this orientation fits relatively well the volume of the reconstruction. Because of radiation damage on the small 2D crystals due to the high electron dose used for the EM procedure, which affected the reconstructed volume, no attempt was made to optimize the fit of the model other than by visual inspection, guided by the absence of clashes with the constraints indicated above. The contacts involve residues in loops BC and C'D of domain III.

Illustrations

All the figures presented, including those in Supplementary Information, were prepared using the program RIBBONS⁴².

Received 25 August; accepted 14 November 2003; doi:10.1038/nature02239.

- Jahn, R., Lang, T. & Sudhof, T. C. Membrane fusion. *Cell* **112**, 519–533 (2003).
- Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* **69**, 531–569 (2000).
- Weissenhorn, W. *et al.* Structural basis for membrane fusion by enveloped viruses. *Mol. Membr. Biol.* **16**, 3–9 (1999).
- Danieli, T., Pelletier, S. L., Henis, Y. I. & White, J. M. Membrane fusion mediated by the influenza virus hemagglutinin requires the concerted action of at least three hemagglutinin trimers. *J. Cell Biol.* **133**, 559–569 (1996).
- Blumenthal, R., Sarkar, D. P., Durell, S., Howard, D. E. & Morris, S. J. Dilation of the influenza hemagglutinin fusion pore revealed by the kinetics of individual cell–cell fusion events. *J. Cell Biol.* **135**, 63–71 (1996).
- Markovic, I., Leikina, E., Zhukovsky, M., Zimmerberg, J. & Chernomordik, L. V. Synchronized activation and refolding of influenza hemagglutinin in multimeric fusion machines. *J. Cell Biol.* **155**, 833–844 (2001).
- Zimmerberg, J. & Chernomordik, L. V. Membrane fusion. *Adv. Drug Deliv. Rev.* **38**, 197–205 (1999).
- Blumenthal, R., Clague, M. J., Durell, S. R. & Epand, R. M. Membrane fusion. *Chem. Rev.* **103**, 53–69 (2003).
- Helenius, A., Kartenbeck, J., Simons, K. & Fries, E. On the entry of Semliki Forest virus into BHK-21 cells. *J. Cell Biol.* **84**, 404–420 (1980).
- Schlesinger, S. & Schlesinger, M. J. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 895–916 (Lippincott Williams and Wilkins, Philadelphia, 2001).
- Lescar, J. *et al.* The fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* **105**, 137–148 (2001).
- Zhang, W. *et al.* Placement of the structural proteins in Sindbis virus. *J. Virol.* **76**, 11645–11658 (2002).
- Rey, F. A., Heinz, F. X., Mandl, C., Kunz, C. & Harrison, S. C. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* **375**, 291–298 (1995).
- Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc. Natl Acad. Sci. USA* **100**, 6899–6901 (2003).
- Wahlberg, J. M. & Garoff, H. Membrane fusion process of Semliki Forest virus I. Low pH-induced rearrangement in spike protein quaternary structure precedes virus penetration into cells. *J. Cell Biol.* **116**, 339–348 (1992).
- Kielian, M., Chatterjee, P. K., Gibbons, D. L. & Lu, Y. E. in *Subcellular Biochemistry* Vol. 34 *Fusion of Biological Membranes and Related Problems* (eds Hilderson, H. & Fuller, S.) 409–455 (Plenum, New York, 2000).
- Klimjack, M. R., Jeffrey, S. & Kielian, M. Membrane and protein interactions of a soluble form of the Semliki Forest virus fusion protein. *J. Virol.* **68**, 6940–6946 (1994).
- Gibbons, D. L. & Kielian, M. Molecular dissection of the Semliki Forest virus homotrimer reveals two functionally distinct regions of the fusion protein. *J. Virol.* **76**, 1194–1205 (2002).

- Ahn, A., Gibbons, D. L. & Kielian, M. The fusion peptide of Semliki Forest virus associates with sterol-rich membrane domains. *J. Virol.* **76**, 3267–3275 (2002).
- Gibbons, D. L. *et al.* Visualization of the target-membrane-inserted fusion protein of Semliki Forest virus by combined electron microscopy and crystallography. *Cell* **114**, 573–583 (2003).
- Eckert, D. M. & Kim, P. S. Mechanisms of viral membrane fusion and its inhibition. *Annu. Rev. Biochem.* **70**, 777–810 (2001).
- Wahlberg, J. M., Bron, R., Wilschut, J. & Garoff, H. Membrane fusion of Semliki Forest virus involves homotrimers of the fusion protein. *J. Virol.* **66**, 7309–7318 (1992).
- Markosyan, R. M., Cohen, F. S. & Melikyan, G. B. HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation. *Mol. Biol. Cell* **14**, 926–938 (2003).
- Kemble, G. W., Danieli, T. & White, J. M. Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. *Cell* **76**, 383–391 (1994).
- Armstrong, R. T., Kushnir, A. S. & White, J. M. The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition. *J. Cell Biol.* **151**, 425–437 (2000).
- Bagai, S. & Lamb, R. A. Truncation of the COOH-terminal region of the paramyxovirus SV5 fusion protein leads to hemifusion but not complete fusion. *J. Cell Biol.* **135**, 73–84 (1996).
- Januszski, M. M., Cannon, P. M., Chen, D., Rozenberg, Y. & Anderson, W. F. Functional analysis of the cytoplasmic tail of Moloney murine leukemia virus envelope protein. *J. Virol.* **71**, 3613–3619 (1997).
- Melikyan, G. B., Markosyan, R. M., Brener, S. A., Rozenberg, Y. & Cohen, F. S. Role of the cytoplasmic tail of ecotropic moloney murine leukemia virus Env protein in fusion pore formation. *J. Virol.* **74**, 447–455 (2000).
- Chernomordik, L., Frolov, V. A., Leikina, E., Bronk, P. & Zimmerberg, J. The pathway of membrane fusion catalyzed by influenza hemagglutinin: restriction of lipids, hemifusion, and lipidic fusion pore formation. *J. Cell Biol.* **140**, 1369–1382 (1998).
- Gaudin, Y. Rabies virus-induced membrane fusion pathway. *J. Cell Biol.* **150**, 601–612 (2000).
- Markosyan, R. M., Melikyan, G. B. & Cohen, F. S. Evolution of intermediates of influenza virus hemagglutinin-mediated fusion revealed by kinetic measurements of pore formation. *Biophys. J.* **80**, 812–821 (2001).
- Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**, 313–319 (2004).
- Bressanelli, S. *et al.* Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *EMBO J.* (in the press).
- Gibbons, D. L. *et al.* Purification and crystallization reveal two types of interactions of the fusion protein homotrimer of Semliki Forest virus. *J. Virol.* (in the press).
- Otwinski, Z. & Minor, W. in *Macromolecular Crystallography Part A* (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic Press, London, 1997).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–493 (1997).
- Cowan, K. dm: an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38 (1994).
- Jones, T. A. & Kjeldgaard, M. in *Macromolecular Crystallography Part B* (eds Carter, C. W. & Sweet, R. M.) 173–208 (Academic Press, London, 1997).
- Roussel, A. & Cambillaud, C. *Silicon Graphics Geometry Partners Directory* (Silicon Graphics, Mountain View, California, 1991).
- Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Carson, M. Ribbon models of macromolecules. *J. Mol. Graph.* **5**, 103–106 (1987).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Bressanelli, S. Duquerry and P. Fernandez Varela for their help at different stages of this work; A. Ahn and A. Urian for help with virus and protein preparation; C. Schulze-Briese and T. Tomikazi for help during diffraction data collection; Y. Gaudin for critically reading the manuscript; and J. Navaza for helpful discussions. More than 80% of the data used to determine the crystal structure were collected at synchrotron beam line X06SA of the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. Other synchrotron sources used were beam lines ID14 and ID29 at the European Synchrotron Radiation Facility, Grenoble, France, and beam line BW7A at DESY, Hamburg, Germany. M.K. acknowledges support from the Public Health Service and from a Cancer Center Core Support Grant from the National Cancer Institute. F.A.R. acknowledges support from the CNRS and INRA, the SESAME Program of the Région Ile-de-France, the French Fondation pour la Recherche Médicale, the Association pour la Recherche contre le Cancer, the CNRS programs “Physique et Chimie du Vivant” and “Dynamique et réactivité des assemblages biologiques”, and the European Union ENHcV consortium. D.L.G. was supported by the Medical Scientist Training Program of the Albert Einstein College of Medicine, the Albert Cass Traveling Fellowship and the CNRS.

Authors' contributions The crystallographic analyses reported in this paper were performed by F.A.R. and M.C.V.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.A.R. (rey@vms.cnrs-gif.fr) or M.K. (kielian@acem.com.yu.edu). The atomic coordinates of the E1* trimer, as well as the observed structure factors, have been deposited in the Protein Data Bank under accession code 1RER.

A distance of 133–137 parsecs to the Pleiades star cluster

Xiaopei Pan¹, M. Shao¹ & S. R. Kulkarni²

¹Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA

²Caltech Optical Observatories 105-24, California Institute of Technology, Pasadena, California 91125, USA

Nearby 'open' clusters of stars (those that are not gravitationally bound) have played a crucial role in the development of stellar astronomy because, as a consequence of the stars having a common age, they provide excellent natural laboratories to test theoretical stellar models. Clusters also play a fundamental part in determining distance scales. The satellite Hipparcos¹ surprisingly found that an extensively studied open cluster—the Pleiades (also known as the Seven Sisters)—had a distance of $D = 118 \pm 4$ pc (refs 2, 3), about ten per cent smaller than the accepted value^{4–6}. The discrepancy generated a spirited debate because the implication⁷ was that either current stellar models were incorrect by a surprising amount or Hipparcos was giving incorrect distances. Here we report the orbital parameters of the bright double star Atlas in the Pleiades, using long-baseline optical/infrared interferometry. From the data we derive a firm lower bound of $D > 127$ pc, with the most likely range being $133 < D < 137$ pc. Our result reaffirms the fidelity of current stellar models.

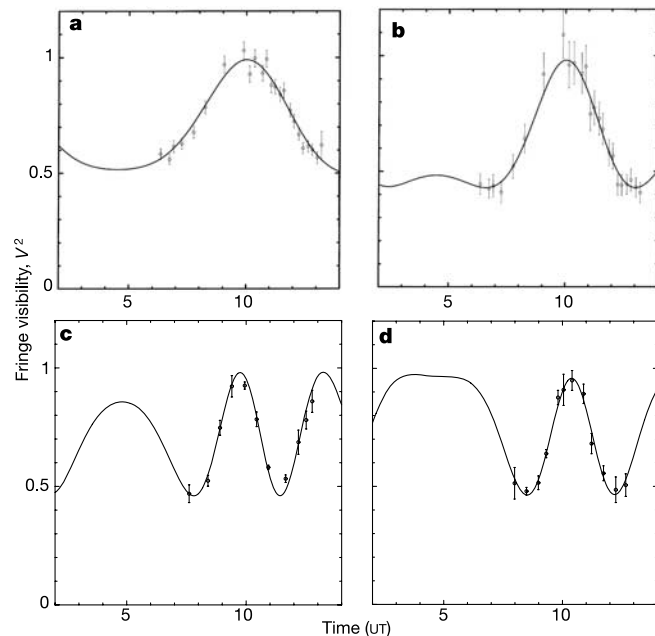


Figure 1 Fringe visibilities and model fits for Atlas. **a**, V^2 measured with the Mark III interferometer (marked by points with 1σ error bars) on UT 6 October 1992 at a mean wavelength of $\bar{\lambda} = 800$ nm. The best-fit model shown by the continuous curve corresponds to a binary with $S_\alpha = 5.07$ mas and $S_\delta = -0.46$ mas and component intensity ratio $r = 0.18$. **b**, As for **a** but at $\bar{\lambda} = 550$ nm. Model parameters are $S_\alpha = 4.96$ mas and $S_\delta = -0.43$ mas and $r = 0.202$. **c**, V^2 measurements from PTI obtained in the $2.2\text{-}\mu\text{m}$ band on UT 11 October 1998 (**c**) and UT 18 October 1998 (**d**). The corresponding model parameters are $S_\alpha = -6.83$ mas, $S_\delta = 7.77$ mas and $r = 0.18$ (for **c**) and $S_\alpha = -6.50$ mas, $S_\delta = 6.25$ mas and $r = 0.18$ (for **d**). The difference in magnitude, averaged over all PTI observations, is $\Delta K = 1.86 \pm 0.06$ and that for the Mark III data is $\Delta V = 1.68 \pm 0.07$. The errors were obtained by propagating the measurement errors.

Atlas (also known as HR1178 and HD23850) is a spectral type B8 giant star (luminosity class III) and, with visual magnitude $V = 3.62$ mag, is the second brightest star in the Pleiades. Although observed since 1903 (refs 8, 9), weak and broad spectral features made it difficult to demonstrate firmly that Atlas is a binary star. In 1974, lunar occultation observations¹⁰ conclusively demonstrated its binary nature. However, given the limited occultation opportunities, the technique is not suitable for determining the orbital parameters.

During 1989–1992 we observed Atlas with the Mark III stellar interferometer¹¹ at optical wavelengths and easily resolved the binary system. Subsequently, between 1996–1999, observations in an infrared band were conducted with the Palomar Testbed Interferometer¹² (PTI). In Fig. 1 we show four nightly measurements of Atlas obtained from the Mark III interferometer and PTI. Each epoch was fitted to a binary model and the angular separations of the primary from the secondary are given in Table 1.

The separations obtained from our interferometric observations (Table 1) were fitted to a keplerian (elliptical) orbit and the resulting orbit (and the orbital parameters) is shown in Fig. 2. The lunar occultation observations^{10,13,14} provide an independent check of our interferometric orbit. Thanks to the high precision of our orbital parameters, in particular the orbital period P_b , we can calculate the stellar separation for the epochs of the occultations over the past three decades. As can be seen from Table 2, the agreement, with one exception, is excellent.

The satellite Hipparcos can, in principle, detect the motion of the photocentre of a binary star with respect to other stars. The photocentric and visual orbits are related as follows:

$$a''_{\text{photo}} = a'' \times \frac{(q-r)}{(1+q)(1+r)} \quad (1)$$

where a''_{photo} and a'' (see Fig. 2) are the semi-major axis of the photocentric orbit and the visual orbit; q and r are the mass and intensity ratio of the secondary to that of the primary.

Table 1 The angular separations in right ascension and declination between the primary and the secondary components

JD 2,400,000+	S_α (mas)	$(M-C)_\alpha$ (mas)	S_δ (mas)	$(M-C)_\delta$ (mas)
Mark III Stellar Interferometer				
47872.7767	-7.50 ± 0.95	-0.34	12.1 ± 0.45	-0.33
48561.9051	4.20 ± 0.65	-0.17	-9.0 ± 0.15	0.02
48900.9468	4.68 ± 0.65	-0.11	-0.69 ± 0.15	-0.00
48901.9339	4.96 ± 0.65	0.24	-0.43 ± 0.15	0.02
48902.9230	4.44 ± 0.65	-0.20	-0.17 ± 0.15	0.04
Palomar Testbed Interferometer				
50359.8709	4.54 ± 0.63	0.12	0.38 ± 0.62	-0.11
50412.8662	-1.03 ± 0.63	0.20	11.05 ± 0.62	-0.26
50420.7811	-2.50 ± 0.63	-0.38	12.40 ± 0.62	0.05
51080.9480	-7.17 ± 0.58	0.04	10.96 ± 0.17	-0.13
51096.9303	-6.63 ± 0.55	0.15	7.84 ± 0.14	-0.17
51097.9347	-6.83 ± 0.51	-0.10	7.77 ± 0.14	-0.01
51104.9298	-6.50 ± 0.51	-0.20	6.25 ± 0.14	0.12
51105.8995	-6.35 ± 0.56	-0.12	6.03 ± 0.16	0.14
51127.8705	-3.93 ± 0.53	-0.10	-0.07 ± 0.14	0.04
51130.8410	-3.56 ± 0.54	-0.16	-0.83 ± 0.16	0.13
51154.7422	1.14 ± 0.53	0.46	-7.16 ± 0.16	-0.21
51155.7434	0.89 ± 0.54	0.04	-7.20 ± 0.16	-0.06
51470.8760	4.03 ± 0.51	-0.44	-8.93 ± 0.14	0.06
51480.9014	5.0 ± 0.51	-0.26	-8.31 ± 0.14	-0.11

S_α and S_δ were obtained by fitting the square of the fringe visibility function, V^2 , to a binary model. See refs 26, 27 for examples of modelling of binary stars with interferometric data. For observations made with the Mark III stellar interferometer, the length of the baseline varied from 15.1 to 27.6 m and the visibility measurements were made in bands centred on 0.8, 0.55, 0.50 or 0.45 μm . For observations made with the Palomar Testbed Interferometer, the physical baseline was 109.8 m long and multiple channels centred on 2.2 μm were used. For these baselines, the inferred binary parameters are insensitive to the angular diameters of the components (which, on the basis of stellar theory, we set to 0.6 mas and 0.4 mas for Atlas A and Atlas B respectively). The separations were used to construct the visual orbit shown in Fig. 2. It is important to note that V^2 analysis yields the separation vector, but with an ambiguity of 180° in the position angle. We have chosen between the two orientations by trial and error, driven by the desire that the orbit smoothly connect between successive epochs. For each axis, the difference between the measurements (M) and the calculated or model (C) is shown in the column $M - C$.

We find the difference in V-band magnitudes of the two components is $\Delta V = 1.68 \pm 0.07$ (Fig. 1) which corresponds to $r = 0.212 \pm 0.014$. We obtain $a''_{\text{photo}} = 0.32a''$ (for $q = 1$) and $a''_{\text{photo}} = 0.25a''$ (for $q = 0.74 \pm 0.05$; see Fig. 3). Thus, a''_{photo} ranges from 4.21 to 3.24 milliarcseconds (mas). The small photocentric orbit is a challenging measurement for Hipparcos—as confirmed by the marginal (4σ) measurement, $a''_{\text{photo}} = 4.23 \pm 0.97$ mas (ref. 1) as well as the non-detection of eccentricity. However, thanks to the better temporal sampling of the Hipparcos data, the orbital period of 290.7 ± 8.6 d is well-determined, and is consistent with our independently derived period (see Fig. 2).

Kepler's third law relates the total mass of the binary system, M_{tot} , to a , the physical semi-major axis: $a^3 \propto M_{\text{tot}} P_b^2$. Noting that $a'' = a/D$, where D is the distance, simple error propagation of Kepler's third law yields:

$$V_R(D) = 1/9 \times V_R(M_{\text{tot}}) + 4/9 \times V_R(P_b) + V_R(a'') \quad (2)$$

Here, $V_R(x) \equiv (\sigma(x)/X)^2$ is the fractional variance of x , where X is the mean value and $\sigma^2(x)$ is the variance in x . As can be seen from Fig. 2, the fractional variance of P_b and a'' is 10^{-7} and 10^{-4} , respectively. Thus, even if the uncertainty in M_{tot} is 10%, given the high precision of our orbital parameters, the uncertainty on D will only be 3%.

To this end, we assumed the Hipparcos parallax and the isochrones of ref. 15 and derived the masses of the primary (M_A) and secondary (M_B) components. The mass estimates in conjunction with Kepler's third law yielded a new distance estimate. The procedure was repeated until the derived distance was the same as the assumed distance. The convergence was swift, requiring only three iterations. The convergence is rapid for the following reason. The mass is derived from the assumed luminosity ($L \propto D^2$) and the mass–luminosity relation, $L \propto M^\beta$; on the main sequence, $\beta \approx 3.8$. Combining with Kepler's third law we see that the derived distance $\propto M^{1/3} \propto D^{2/(3\beta)}$ is weakly dependent on the assumed distance.

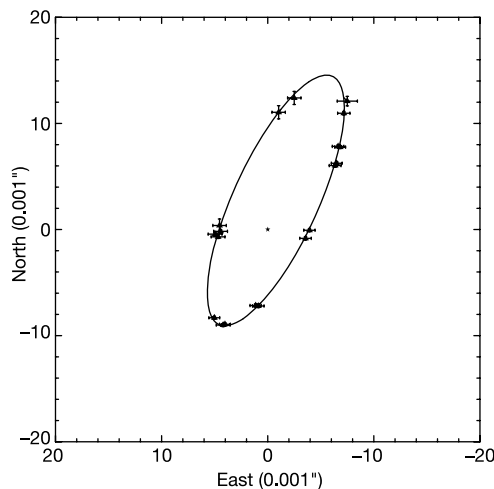


Figure 2 The visual orbit of Atlas. Each of the separation vectors displayed in Table 1 is represented by a triangle and is connected to the calculated positions by a short line. The symbol * refers to the primary star. The crosses indicate the formal 1σ errors (Table 1). A keplerian (elliptical) model was fitted to the separation vector in the framework of differential corrections (see ref. 26) to a binary orbit. The orbital elements of Atlas obtained from our observations are as follows: orbital period, $P_b = 290.81 \pm 0.06$ d; epoch, $T_0 = \text{JD } 2450582.8 \pm 1.5$ d; eccentricity, $e = 0.2457 \pm 0.006$; semi-major axis of the ellipse, $a' = 12.94 \pm 0.11$ mas; and the three orientation angles, $i = 108.5^\circ \pm 0.5^\circ$, $\omega = 332.1^\circ \pm 1.4^\circ$ and $\Omega = 155.0^\circ \pm 0.6^\circ$. As an additional check on our orbital period, P_b (specifically the concern over ambiguities due to missed cycles), we explored the range 143 d to 590 d of orbital period and find the signal is strongest at $P_b = 290.81 \pm 0.06$ d.

We find $M_B = 3.57M_\odot$ (for $\tau_p = 120$ Myr; here τ_p is the age of the Pleiades cluster) or $M_B = 3.72M_\odot$ (for $\tau_p = 90$ Myr) and $D = 136$ pc and an uncertainty of 2.6 pc arising from the uncertainty in τ_p . A strong lower limit of 127 pc is obtained by making the assumption $M_{\text{tot}} > 2M_B$. The results are summarized in Fig. 3. The uncertainty in D is comparable to the half-mass radius (of about 1.5 pc; refs 15, 16) for evolved stars in the Pleiades. Given that Atlas is either the second or third massive star in Pleiades it is likely that Atlas is in the core and not in the extended halo of Pleiades.

From our model we expect radial-velocity amplitudes of $K_A = 27 \text{ km s}^{-1}$ and $K_B = 36 \text{ km s}^{-1}$. With modern spectrographs it may be possible to obtain radial-velocity data with a precision of 1 km s^{-1} (preliminary observations by P. North, Ecole Polytechnique Fédérale de Lausanne). Such data, when combined with our interferometric model, can yield a distance estimate to the Pleiades that is accurate to 2%, independent of stellar models.

Our result reaffirms the traditional distance estimate^{4–6} to the Pleiades and thus gives renewed confidence to the basic stellar model formulation used by astronomers. In particular, changing¹⁸ the abundance of metals or that of He may result in a brighter main sequence but as noted above in equation (2), small changes in M_{tot} will result in even smaller changes in D .

Next, our parallax for Atlas is in disagreement with that obtained by Hipparcos. The mean Hipparcos parallax for the brightest (that is, the most massive) fifteen stars is 8.68 ± 0.21 mas, translating to a distance range of 107–124 pc (3σ). Thus we conclude that Hipparcos parallaxes for the Pleiades apparently suffer from an additional error of about 1 mas. The results presented here lend credence to the hypothesis¹⁹ that Hipparcos measurements may be contaminated by systematic errors on small angular scales. Because such a conclusion would have important implications for Hipparcos parallaxes, it is imperative that accurate distances be derived for several other stars in the Pleiades.

As demonstrated by the work reported here, observations of a single binary can result in both an accurate and precise distance to the cluster. Interferometric observations, especially when combined with radial-velocity measurements, yield distance estimates that are independent of stellar models and extinction. With such distance estimates, stellar astronomers can focus on using open clusters as natural ‘laboratories’ in which to test their models and probe subtle effects due to rotation, stellar activity and convection.

Specialized modes of interferometry, in particular phase closure²⁰ and very-narrow-angle astrometry²¹, offer even-higher-precision orbit determinations. Interferometry combined with precision radial-velocity measurements (particularly double-lined binaries²²) can be used to test stellar models rigorously, given that both

Table 2 Comparison of the interferometric model with lunar occultation observations

JD 2,400,000+	Occultation		Interferometry		Calculated ρ_p (mas)	$\rho_1 - \rho_p$ (mas)	Ref.
	$\phi_1(^{\circ})$	ρ_1 (mas)	ρ (mas)	$\theta(^{\circ})$			
40221.6	56	6.2	9.23	181.6	5.4	0.8	10
41396.4	124	7.4	11.17	174.5	7.1	0.3	10
41724.7	49	4.0	15.28	161.2	5.8	–1.8	10
41397.1	93.7	2.5	11.27	–5.9	1.9	0.6	13
46988.2	21.1	6.8	15.29	–26.5	10.3	–3.5	14
48307.8	294.4	6.2	6.43	122.2	6.4	–0.2	14

JD, the epoch (Julian Day) of the lunar occultations. ϕ_1 is the position angle of the lunar limb at the time of occultation. ρ_1 is the projected binary separation inferred from the occultation observations. The polar coordinates of the separation between the primary and secondary components extrapolated from the binary model presented in Fig. 2 are ρ and θ (the angle measured from North towards East). Our orbital model has precision sufficient to extrapolate from the 1990s to the period 1968–1991 over which the lunar occultation observations were conducted. The calculated projected separation is $\rho_p = \rho \cos(\theta - \phi_1)$. Our reading of the literature suggests that the typical error in lunar occultation measurements is about 1 mas. As can be seen from the last column, our model can account for the lunar occultation observations within errors. A measurement on 29 December 1971 (ref. 13) was not used, given the cautionary statements made by the authors. Lunar occultation experiments¹⁴ also yield the magnitude difference: $\Delta V = 1.6 \pm 0.2$ mag and $\Delta K = 1.80 \pm 0.21$ mag. These are consistent with our results in Fig. 1.

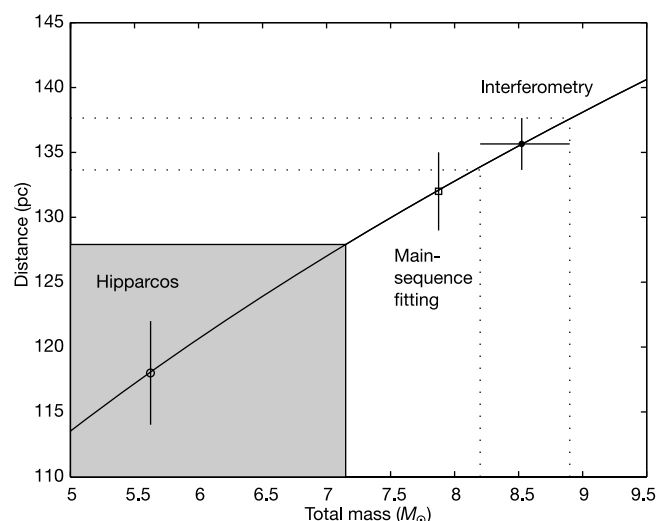


Figure 3 Mass–distance relationship for Atlas. Using the measured angular length of a'' and P_b (see Fig. 2), Kepler's third law allows us to relate the distance D to $M_{\text{tot}} = M_A + M_B$ of the binary system (black curve). The final uncertainty in D is dominated by uncertainty in M_{tot} (see below) and for this reason the black curve is shown without the width arising from uncertainties in the orbital parameters. We derive the masses of the components by comparing their observed V magnitudes (Fig. 1) against the stellar models of ref. 15. Here and throughout the text, following ref. 4, we assume a mean interstellar extinction of $E_{B-V} = 0.04$ mag. We find $M_B = 3.57 M_{\odot}$ (for assumed Pleiades cluster age, $\tau_P = 120$ Myr) or $M_B = 3.72 M_{\odot}$ (for $\tau_P = 90$ Myr). The mass of the secondary (apart from the uncertainty in the assumed age) is robust because on the main sequence the luminosity increases steeply with mass ($L \propto M^{\beta}$ with $\beta \approx 3.8$) and this relation, when combined with Kepler's third law, makes the derived distance almost independent of the assumed mass (see the discussion following equation (2)). The star Maia, presumably a single star with $V = 3.87$, is a twin of Atlas A but with well-determined colours. Using isochrones that allow for convective overshoot¹⁵, estimates for M_A vary from $4.62 M_{\odot}$ (for $\tau_P = 120$ Myr) to $5.19 M_{\odot}$ (for $\tau_P = 90$ Myr); isochrones supplied by D. VandenBerg (personal communication) clearly favour the latter age and $M_A = 5.19 M_{\odot}$. Thus, $M_{\text{tot}} = 8.2$ to $8.9 M_{\odot}$, which leads to D being uncertain from 133.9 pc to 137.6 pc (the dotted lines). Black circle, the arithmetic mean of these two estimates. The error bar is ± 2.6 pc, arising from the uncertainty in M_{tot} . Square, the distance to the Pleiades based on main-sequence fitting⁵. White circle, the Hipparcos distance estimate^{2,3} of 118 ± 4 pc. Neither technique involves the masses of the components of Atlas and hence no horizontal error bars are shown. As argued above, at the very minimum, $M_{\text{tot}} > 2M_B > 7.14 M_{\odot}$, which translates to $D > 127$ pc. The shaded region marks the parameter space excluded by our observations.

components are expected to lie on exactly the same isochrone. Radial-velocity studies²³, lunar occultation²⁴ and photometric²⁵ observations have revealed many binaries in the Pleiades. The availability of new-generation high-resolution spectrographs and the recent commissioning of powerful interferometers at the Keck Observatory and the Very Large Telescope facility are expected to facilitate such testing of stellar models. □

Received 29 September; accepted 11 December 2003; doi:10.1038/nature02296.

1. Perryman, M. A. C. *et al.* The HIPPARCOS Catalogue. *Astron. Astrophys.* **323**, L49–L52 (1997).
2. Mermilliod, J.-C., Turon, C., Robichon, N., Arenou, F. & Lebreton, Y. in *ESA SP-402: Hipparcos—Venice'97* (eds Perryman, M. A. C. & Bernacca, P. L.) 643–650 (European Space Agency, Paris, 1997).
3. van Leeuwen, F. & Hansen Ruiz, C. S. in *ESA SP-402: Hipparcos—Venice'97* (eds Perryman, M. A. C. & Bernacca, P. L.) 689–692 (European Space Agency, Paris, 1997).
4. Meynet, G., Mermilliod, J.-C. & Maeder, A. A New dating of galactic open clusters. *Astron. Astrophys. Suppl. Ser.* **98**, 477–504 (1993).
5. Pinsonneault, M. H., Stauffer, J., Soderblom, D. R., King, J. R. & Hanson, R. B. The problem of HIPPARCOS distances to open clusters. I. Constraints from multicolor main-sequence fitting. *Astrophys. J.* **504**, 170–191 (1998).
6. Gatewood, G., de Jonge, J. K. & Han, I. The Pleiades, MAP-based trigonometric parallaxes of open clusters. *V. Astrophys. J.* **533**, 938–943 (2000).
7. Paczynski, B. The distance to Pleiades. *Acta Astron.* **53**, 209–211 (2003).
8. Abt, H. A., Barnes, R. C., Biggs, E. S. & Osmer, P. S. The frequency of spectroscopic binaries in the Pleiades. *Astrophys. J.* **142**, 1604–1615 (1965).

9. Pearce, J. A. & Hill, G. Four suspected spectroscopic binaries in the Pleiades. *Publ. Astron. Soc. Pacif.* **83**, 493–495 (1971).
10. McGraw, J. T., Dunham, D. W., Evans, D. S. & Moffet, T. J. Occultations of the Pleiades: photoelectric observations at Tonantzintla with a discussion of the duplicity of Atlas. *Astron. J.* **79**, 1299–1303 (1974).
11. Shao, M., Colavita, M. M., Hines, B. E., Staelin, D. H. & Hutter, D. J. The Mark III stellar interferometer. *Astron. Astrophys.* **193**, 357–371 (1988).
12. Colavita, M. M. *et al.* The Palomar testbed interferometer. *Astrophys. J.* **510**, 505–521 (1999).
13. de Vegt, C. & Gehlich, U. K. Results of photoelectric lunar occultation observations obtained at the Hamburg Observatory during 1969–1973. *Astron. Astrophys.* **48**, 245–252 (1976).
14. Meyer, C., Rabbia, Y., Froeschle, M., Helmer, G. & Amieux, G. Observations of lunar occultations at Observatoire de la Côte d'Azur. *Astron. Astrophys. Suppl. Ser.* **110**, 107–123 (1995).
15. Ventura, P., Zepieri, A., Mazzitelli, I. & D'Antona, F. Full spectrum of turbulence convective mixing: I. Theoretical main sequences and turn-off for 0.6–15 $M_{\odot}$. *Astron. Astrophys.* **334**, 953–968 (1998).
16. Raboud, D. & Mermilliod, J.-C. Investigation of the Pleiades cluster. IV. The radial structure. *Astron. Astrophys.* **329**, 101–114 (1998).
17. Pinfield, D. J., Jameson, R. F. & Hodgkin, S. T. The mass of the Pleiades. *Mon. Not. R. Astron. Soc.* **299**, 955–964 (1998).
18. Castellani, V., Degl'Innocenti, S., Prada Moroni, P. G. & Tordiglione, V. Hipparcos open clusters and stellar evolution. *Mon. Not. R. Astron. Soc.* **334**, 193–197 (2002).
19. Narayanan, V. K. & Gould, A. Correlated errors in HIPPARCOS parallaxes toward the Pleiades and the Hyades. *Astrophys. J.* **523**, 328–339 (1999).
20. Shao, M. & Colavita, M. M. Long-baseline optical and infrared stellar interferometry. *Annu. Rev. Astron. Astrophys.* **30**, 457–498 (1992).
21. Lane, B. F. & Mutterspaugh, M. W. Differential astrometry of sub-arcsecond scale binaries at the Palomar testbed interferometer. *Astron. J.* **125**, 1623–1628 (2003).
22. Giannuzzi, M. A. The spectroscopic binary HD 23642 and the distance of the Pleiades. *Astron. Astrophys.* **293**, 360–362 (1995).
23. Mermilliod, J.-C., Rosvick, J. M., Duquenois, A. & Mayor, M. Investigation of the Pleiades cluster. II—Binary stars in the F5–K0 spectral region. *Astron. Astrophys.* **265**, 513–526 (1992).
24. Richichi, A., Calamai, G. & Leinert, C. New binary stars discovered by lunar occultations. *Astron. Astrophys.* **286**, 829–837 (1994).
25. Torres, G. Discovery of a bright eclipsing binary in the Pleiades cluster. *Inform. Bull. Variable Stars* **5402**, 1–3 (2003).
26. Pan, X. P. *et al.* Apparent orbit of the spectroscopic binary β Arietis with the Mark III stellar interferometer. *Astrophys. J.* **356**, 641–645 (1990).
27. Hummel, C. A. *et al.* The spectroscopic binary ϵ Andromedae: Determination of the orbit by optical interferometry. *Astron. J.* **106**, 2486–2492 (1993).

Acknowledgements We thank other members of the PTI and Mark III teams and especially A. Boden and B. Lane for a careful reading. We gratefully acknowledge discussions with N. Reid, R. M. Rich, D. Sasselov, J. Stauffer, J. Tomkin and D. VandenBerg. We extensively used the SIMBAD database of astronomical papers and are grateful to CDS, France, and NASA for maintaining this system. The research described in this paper was primarily carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. S.R.K.'s research is supported by NSF and NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.R.K. (srk@anjucaltech.edu).

The microscopic nature of localization in the quantum Hall effect

S. Ilani¹*, J. Martin¹, E. Teitelbaum¹, J. H. Smet², D. Mahalu¹, V. Umansky¹ & A. Yacoby¹

¹Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel

²Max-Planck-Institut für Festkörperforschung, D-70569 Stuttgart, Germany

* Present address: Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA

The quantum Hall effect arises from the interplay between localized and extended states that form when electrons, confined to two dimensions, are subject to a perpendicular magnetic field¹. The effect involves exact quantization of all the electronic transport properties owing to particle localization. In the conventional theory of the quantum Hall effect, strong-field localization is associated with a single-particle drift motion of electrons along

contours of constant disorder potential². Transport experiments that probe the extended states in the transition regions between quantum Hall phases have been used to test both the theory and its implications for quantum Hall phase transitions. Although several experiments^{3–9} on highly disordered samples have affirmed the validity of the single-particle picture, other experiments^{10–12} and some recent theories^{13–15} have found deviations from the predicted universal behaviour. Here we use a scanning single-electron transistor to probe the individual localized states, which we find to be strikingly different from the predictions of single-particle theory. The states are mainly determined by Coulomb interactions, and appear only when quantization of kinetic energy limits the screening ability of electrons. We conclude that the quantum Hall effect has a greater diversity of regimes and phase transitions than predicted by the single-particle framework. Our experiments suggest a unified picture of localization in which the single-particle model is valid only in the limit of strong disorder.

Unlike extended electrons, whose charge is spread over the entire volume of the system, a localized electron is confined to a small region in space, determined by its localization length. Upon population, a single electronic charge enters this confined region and results in a discrete jump in the local chemical potential μ and a spike in its derivative with respect to the back-gate voltage $d\mu/dV_{BG}$. Hence, the local derivative $d\mu/dV_{BG}$, which is inversely proportional to the local electronic compressibility¹⁶, provides direct

information about the localized states within quantum Hall (QH) phases. To measure this derivative, we use a single-electron transistor (SET) electrometer^{16–20}. Figure 1a shows a coarse measurement of $d\mu/dV_{BG}$ as function of magnetic field (B) and density (n), which is tuned by the back-gate voltage. Incompressible behaviour appears along integer and fractional values of $\nu = hn/eB$ (bright strips) corresponding to integer and fractional QH phases (here h is Planck's constant, and e is the charge on an electron). A detailed measurement of $d\mu/dV_{BG}$ within the QH phases (Fig. 1b) reveals a rich fine structure of parallel (black) lines. Each line is the evolution in the B – n plane of a single spike in $d\mu/dV_{BG}$ and corresponds to charging of an individual localized state.

Figure 1b shows two distinct groups of localized states that surround the $\nu = 1$ and $\nu = 2$ QH phases. Similar groups appear around every well-developed integer and fractional QH phase. In addition, a group of horizontal lines that are associated with the insulating phase ($\nu = 0$) is observed at low densities (Fig. 1a). Further examination of these groups reveals three interesting properties. First, the number of states within each group is independent of B . This is surprising, because, for example, a twofold increase in B results in a twofold increase in the Landau level degeneracy, but hardly changes the number of localized states we observe (Fig. 1c). Second, the groups of localized states form strips of constant width, Δn , in the n – B plane, and Δn is the same for different QH phases (Fig. 1a). In fact, when the longitudinal resistivity (ρ_{xx}) is measured in the same sample, QH behaviour (zero ρ_{xx}) appears also along strips with similar Δn (Fig. 1d). The quantitative correspondence between the transport and local compressibility implies that we probe most or all of the localized states that are responsible for the QH phenomena in transport. Third, localized states that belong to the same group evolve in the n – B plane with exactly the same slope: a slope equal to the slope of their underlying QH phase. Thus, the slope of the charging lines is quantized to $dn/dB = e\nu/h$ (ν integer or fractional). The slope, therefore, identifies each localized state with the QH phase it belongs to, and it remains distinctively quantized even in regions where two groups of states overlap (Fig. 1e). Similar phenomenology is observed in all the samples we have studied (four different samples), which cover a large range of peak mobilities ($(0.5 - 4) \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$).

The measured pattern of localized states is summarized schematically in Fig. 2a. States are bunched in groups that correspond to the various phases of the system. Each group contains a constant number of lines with identical slope. We will now show that these observations are inconsistent with the single-particle description of localized states. Imagine a model disorder potential in the shape of a Mexican hat (Fig. 2a inset). In the single-particle picture, the localized states are constant-energy orbits and enclose integer number of flux quanta. For our model potential, these orbits form concentric circles (blue lines). When B increases, all orbits shrink to maintain constant flux through their area (magenta lines). Thus, at higher B there are more states per unit area. This is a generic property of the single-particle picture, which results from the enhanced Landau level degeneracy as B increases. A microscopic compressibility measurement at any arbitrary location would therefore reveal an increasing number of localized states as the degeneracy of the Landau levels increases, which is in sharp contrast to the constant number of localized states observed. Furthermore, if we evaluate in our simple model the energy of an individual state as function of B (for example, the state shown green) we notice that it is non-monotonic—first going down to the bottom of the potential landscape and subsequently climbing up the central hill. Correspondingly, its charging line in the n – B plane (green line in Fig. 2a) first descends and then rises with a slope that asymptotically approaches a non-quantized value. This non-universal behaviour of the charging lines is another generic property of single-particle states and is in sharp contrast with our observations.

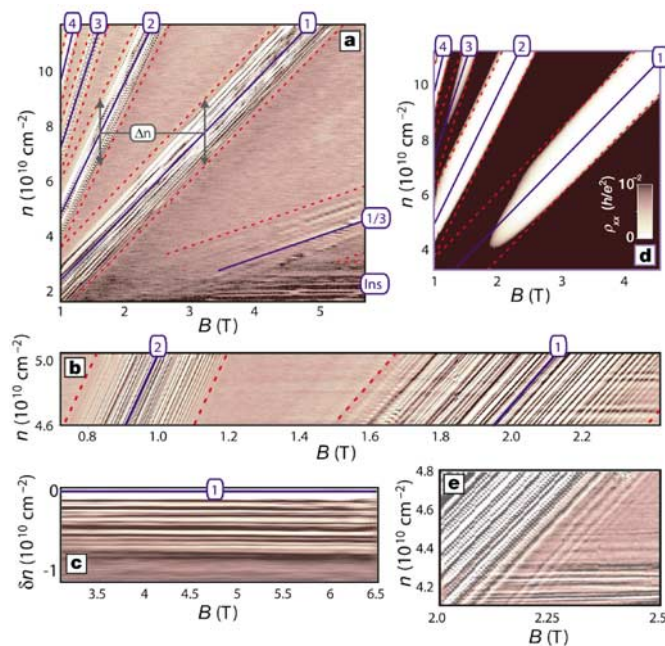


Figure 1 Measurements of the derivative $d\mu/dV_{BG}$ at an arbitrary position above the 2DES as function of magnetic field (B) and density (n). This derivative is measured with a single-electron transistor (SET), which offers a spatial resolution of $\sim 1,000 \text{ \AA}$ and an unprecedented voltage sensitivity of $\sim 1 \text{ } \mu\text{V Hz}^{-1/2}$. **a**, A measurement over a large range in B and n demonstrating the alternating pattern of incompressible (bright) and compressible (dark) regions. The bright regions correspond to the QH phases of the system. **b**, Detailed measurement of the $\nu = 1$ and $\nu = 2$ incompressible regions, revealing a rich fine structure of parallel (black) lines, each representing the charging line of an individual localized state. All charging lines within a certain group have exactly the same slope in the n – B plane. **c**, The charging lines around $\nu = 1$ plotted as function of B and the density deviation from the full Landau level, δn . **d**, The longitudinal resistivity, ρ_{xx} , measured on the same sample as function of B and n . Red lines mark the boundaries, copied from **a**, where localized states disappear. **e**, An overlap region of the $\nu = 1$ and $\nu = 0$ groups showing that localized states of these different groups coexist and maintain their original slopes.

Single-particle physics fails to describe our observations because it ignores the interactions between electrons and hence screening. Electrons in this single-particle picture populate states in a fixed disorder potential. However, Coulomb repulsions produce a potential landscape that varies with the filling factor in an attempt to minimize the total electrostatic energy²¹. Such self-consistent potential landscape is a result of spatial density fluctuations, formed by electrons to oppose the bare disorder potential, $V_{\text{bare}}(r)$, where r is position in real space. For any $V_{\text{bare}}(r)$ a specific density profile exists, $n_{\text{scrn}}(r)$, which screens it completely and thereby minimizes the electrostatic energy. This disorder-induced density profile, which follows from the Poisson equation $\nabla^2 V_{\text{bare}}(x, y, z) = -4\pi en_{\text{scrn}}(x, y)\delta(z)$, has a typical amplitude $\Delta n_{\text{disorder}}$ and is uniquely defined up to an arbitrary constant, which we shall take to be zero. We will show that the evolution of the electronic states is governed mainly by the ability or inability of the electron gas to form this screening density profile.

An intuitive picture of localization induced by interaction is obtained by tracing the self-consistent potential as a function of B and n . Figure 2b depicts the calculated density profiles and the corresponding self-consistent potentials obtained rigorously within the Thomas–Fermi approximation²² for an almost-empty (I), half-full (II), and nearly-full (III) lowest Landau level with typical disorder. For a given B the density in the Landau level cannot exceed one electron per flux quanta, $n_{\text{max}} = B/\phi_0$, and is therefore constrained by $0 \leq n \leq n_{\text{max}}$. At the centre of the Landau level this constraint is insignificant because the variations in density required for screening ($\Delta n_{\text{disorder}}$) are smaller than n_{max} . The density follows the profile required for perfect screening, $n(r) = n_{\text{scrn}}(r) + \langle n \rangle$, where $\langle n \rangle$ is the average density, thus leading to a flat electrostatic potential (II in Fig. 2b). Each electron added to the system experiences this flat potential and thus, within this approximation, is completely delocalized. Hence, as the density increases, $n(r)$ grows uniformly while maintaining its spatial dependence to ensure perfect screening. Electrons fail to flatten the bare disorder potential when the desired $n(r)$ for perfect screening exceeds n_{max} . This commences near the peaks of $n(r)$, leading to the formation of local incompressible regions with a full Landau level ($n(r) = n_{\text{max}}$) where the unscreened bare potential sticks out (III in Fig. 2b). These regions coexist with compressible regions where the Landau level is

not yet filled ($n(r) < n_{\text{max}}$), having a screened potential. Once a compressible region becomes surrounded by an incompressible barrier it behaves as a quantum dot whose charging is governed by Coulomb-blockade physics, that is, electrons are localized to this region and enter it one by one.

The charging spectra of such interaction-induced localized states are very different from those emanating from the single-particle picture, and account for all the properties of localized states we observed. For simplicity we describe in more detail the states around $\nu = 1$. Along the $\nu = 1$ line, where the first Landau level is exactly full, the bare potential is unscreened and the two-dimensional electron system (2DES) is incompressible everywhere. Reducing the average density introduces holes in the Landau level that accumulate at compressible pockets ($\delta n(r) \equiv n(r) - n_{\text{max}} < 0$) surrounded by an incompressible QH phase ($\delta n(r) = 0$). These pockets behave as quantum antidots (or dots for $\delta n(r) > 0$). For a given antidot, holes are added at discrete values of local density deviation, δn_i , and a specific number of holes can be introduced before the confining incompressible barriers disappear. The values δn_i are determined solely by electrostatics and are independent of B . The only effect of B is to increase n_{max} , thereby converting each charging point into a line in the n – B plane. This results in a constant number of charging lines running parallel to the $\nu = 1$ line, in accordance with our measurements. We note that the formation of localized states is determined solely by the density deviation from a completely full Landau level, δn , and not by the total density, n . Hence, despite the increase of Landau level degeneracy (n_{max}) with B , the number of localized states remains constant.

Dots also emerge when electrons start filling an empty Landau level (I in Fig. 2b). Here compressible pockets ($\delta n(r) \equiv n(r) - 0 > 0$) are enclosed by incompressible empty regions ($\delta n(r) = n(r) = 0$), resulting in charging lines that are parallel to $\nu = 0$. Hence, it is the incompressible phase surrounding a dot that determines the quantized slopes of its localized states. In the experiment, we clearly resolve the boundaries where localized states cease to exist. They appear when the compressible dots form a connected network, lose their quantum confinement, and therefore lose their discrete charging spectra. This percolation transition of a compressible medium in an incompressible background^{21,23,24} occurs when the density of compressible fluid equals $\delta n = \Delta n_{\text{disorder}}/2$. Within each

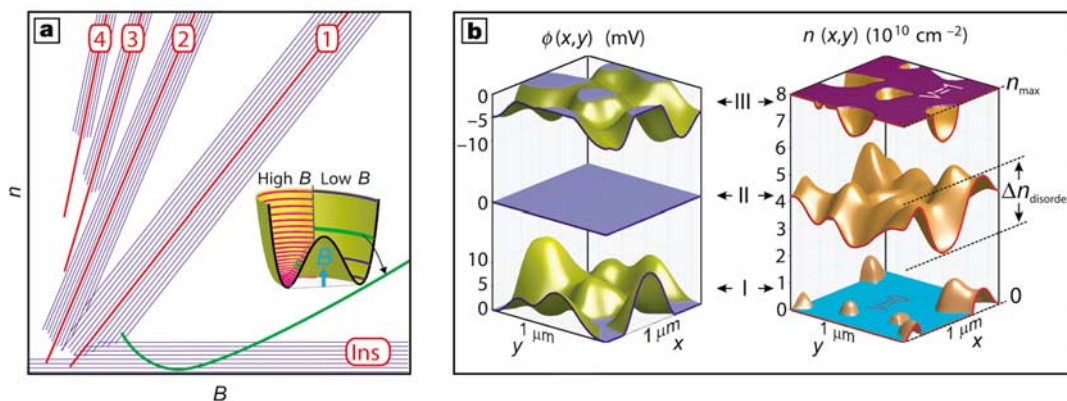


Figure 2 Theoretical models for localized states. **a**, Schematic illustration of the measured charging lines in the n – B plane (blue) compared with a calculated single-particle charging line (green) of a model Mexican hat potential (inset). The single-particle states of the Mexican hat are depicted for low B (blue) and high B (magenta). The specific state that gives rise to the green charging line is marked for both low and high B . Its energy approaches the top of the potential hill in the limit of strong B . This energy is intermediate between the bottom and top of the disorder potential, and therefore leads to a non-quantized asymptotic slope whose value is between $\nu = 0$ and $\nu = 1$. The scatter of potential peak heights in a typical disorder will therefore result in charging lines having

arbitrary slopes in the range $\nu = 0 \rightarrow 1$. **b**, The density profiles, $n(x, y)$, and electrostatic potential profiles, $\phi(x, y)$, calculated self-consistently within the Thomas–Fermi scheme, for an almost-empty (I), half-full (II) and almost-full Landau level (III), in a typical disorder potential. Following ref. 22, the regions in which the electron density is free to fluctuate are taken as infinitely compressible. Near half-filling, the amplitude of the density fluctuations is $\Delta n_{\text{disorder}}$ and the potential is completely screened. Near the bottom of the Landau level ($n = 0$) and its top ($n = n_{\text{max}}$) the potential is screened only in compressible pockets, which are surrounded by incompressible regions, marked as $\nu = 0$ or $\nu = 1$.

Landau level two such percolation transitions take place, one for electrons, when $n = \Delta n_{\text{disorder}}/2$ and one for holes, when $n = n_{\text{max}} - \Delta n_{\text{disorder}}/2$. This explains why localized states and QH transport phenomena appear only in strips around quantized values of ν and associates the width of the strips with $\Delta n_{\text{disorder}}$. The observation that different QH phases display the same width in density is thus a natural consequence of the dot model. The number of localized states within the strips is independent of B . At higher B , new electrons add up to each Landau level owing to its enhanced degeneracy, but these electrons populate non-localized states within the percolating compressible regions, keeping the number of localized states fixed.

Further corroborating evidence for the dot model comes from spatial imaging of the localized states. A SET is mounted on a tip of a scanning microscope (Fig. 3a) and $d\mu/dV_{\text{BG}}$ is measured as function of position (x) and density at temperature $T = 0.3$ K. In the n - x plane each localized state appears as a charging segment, whose length in x represent the spatial extent of the localized state convoluted with the tip response function (Fig. 3b). A spatial scan for the nearly full lowest Landau level is plotted in Fig. 3c. Charging segments can be seen below and above the $\nu = 1$ line. Localized states are not distributed randomly, but rather observed at specific locations with equally spaced charging spectra. These spectra represent generic behaviour observed at all positions across the sample and in different samples, covering a large range of mobilities. They directly manifest the presence of quantum dots—each dot gives rise to a train of Coulomb-blockade peaks at a certain position with level separation given by its charging energy, determined primarily by its area.

In the dot model, the spectra of localized states are determined only by the bare disorder potential and the presence of an energy gap. To verify this prediction we compare in Fig. 3c and d the spectra measured at the same position around gaps of totally different nature: the spin gap at $\nu = 1$ and the orbital gap at $\nu = 2$. The

similarity is evident—both dot and antidot appear in the scans at exactly the same locations, accommodate the same number of electrons and possess the same level spacing.

A different regime is entered when n_{max} becomes smaller than the size of density fluctuations in the 2DES ($\Delta n_{\text{disorder}}$). In any given sample, this occurs for low enough B . In this highly disordered regime, the separation between groups of localized states (n_{max}) drops below their width ($\Delta n_{\text{disorder}}$) and charging lines from different groups start to overlap. Experimentally, rather than merging or bending, lines from different groups maintain their original slope and coexist (Fig. 1e). This coexistence implies that phase mixing might exist on microscopic scales. To capture the microscopic picture in this regime we follow in Fig. 4a the $\nu = 1$ antidot of Fig. 3 down to low magnetic fields. As B decreases its spectrum shifts down with the top of the first Landau level (top density in all figures) while retaining its internal structure. To the left, another spectrum of a $\nu = 0$ dot is visible, which remains intact upon changes in B . The fact that the internal structure of both the $\nu = 0$ dot and the $\nu = 1$ antidot are independent of B demonstrates once more that these spectra are mainly determined by charging energy and not single-particle localization. At the lowest values of B , the dot and antidot coexist at the same electron density and their localized states are less than $1 \mu\text{m}$ apart. Figure 4b shows a self-consistent calculation of $n(r)$ in this highly disordered regime, to clarify that this is indeed plausible. The small degeneracy of the Landau level, n_{max} , chops off the density profile ($0 \leq n \leq n_{\text{max}}$). The sharp jumps from empty to completely full Landau level result in the coexistence of $\nu = 0$ and $\nu = 1$ regions, and compressible pockets in these regions produce $\nu = 0$ dots that neighbour $\nu = 1$ antidots. A new type of compressible strips appear at the boundaries between the $\nu = 0$ and $\nu = 1$ regions^{13–15}. These strips follow the contours of the bare disorder potential (Fig. 4b), reminiscent of single-particle states. In the strong disorder limit, only these compressible strips survive and generate single-particle-like behaviour.

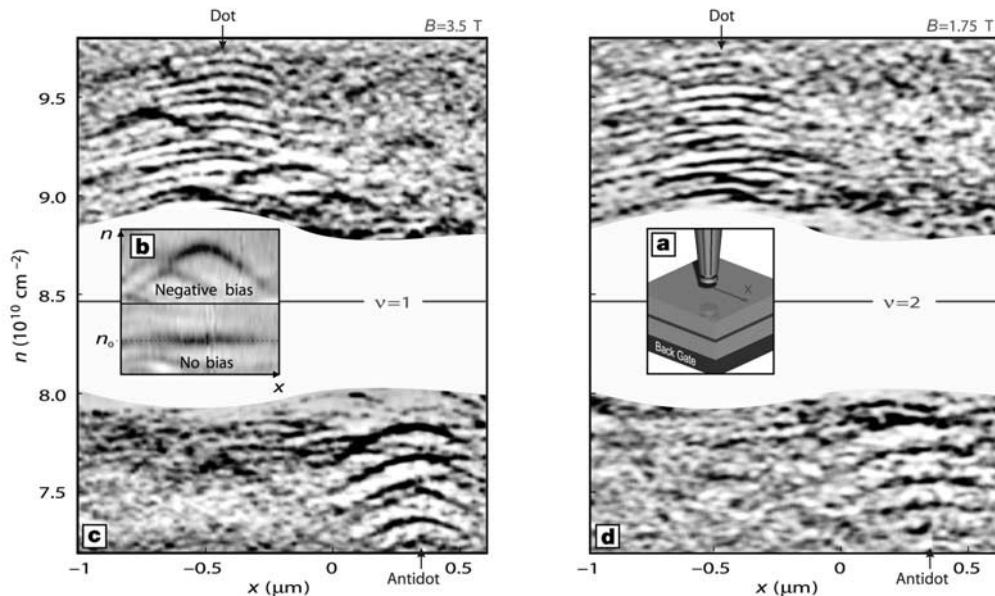


Figure 3 Spatial scans of localized states. **a**, A SET mounted on a tip of a scanning microscope is used to measure $d\mu/dV_{\text{BG}}$ as a function of position, x , and density, n . **b**, The measured signature of a localized state in the n - x plane. The state has a finite spatial extent. Its charging event is thus detected by the SET over a finite range of tip positions. In its non-perturbing state, the tip does not affect the density in which the state is charged (n_0). It results in a straight segment parallel to the x axis, which spans the actual spatial spread of the state convoluted with the tip response function. However, if a slight bias is added to the tip, it acts also as a local gate, pushing the charging condition to higher densities in a manner proportional to the local charge density within the localized state.

This results in a bent segment whose shape gives additional information on the charge distribution within the individual state. **c**, **d**, Spectra of localized states measured around $\nu = 1$ and $\nu = 2$, at the same positions and densities, with a small bias applied to the tip. The localized states bunch spatially into well-ordered families, marked as 'Dot' and 'Antidot'. Despite the difference in the Landau levels and the energy gaps for $\nu = 1$ and $\nu = 2$, their localized states spectra show remarkable similarity. (The central region in both scans is masked, because, as opposed to the rest of the scan, high resistance of the 2DES prevents the system from equilibrating to the thermodynamic ground state.)

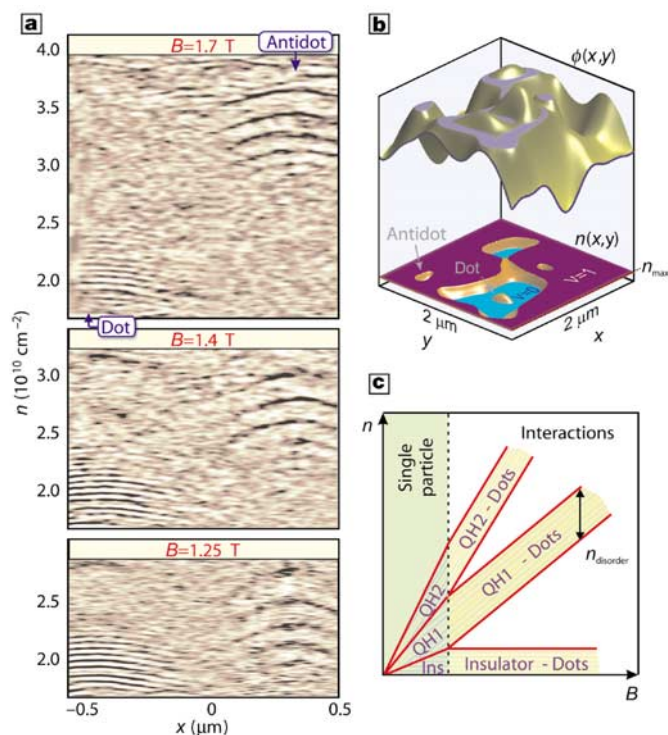


Figure 4 Localized states in the presence of strong disorder. **a**, Density-position scans at $B = 1.7$ T, 1.4 T, 1.25 T tracking a $\nu = 1$ antidot and $\nu = 0$ dot down to the highly disordered regime at low B ($\Delta n_{\text{disorder}} \gg n_{\text{max}}$). The highest density in all scans corresponds to $\nu = 1$. At the lowest B , both dot and antidot coexist—hole-like $\nu = 1$ states neighbour electron-like $\nu = 0$ states less than $1 \mu\text{m}$ away. **b**, Self-consistent Thomas-Fermi calculation in the limit of strong disorder. Here, the disorder creates large density fluctuations, which force the Landau level to be either completely empty or completely full and therefore incompressible. Compressible pockets in the $\nu = 0$ and $\nu = 1$ plateaus result in $\nu = 0$ dots and $\nu = 1$ antidots, as marked in the figure. On the boundary between the $\nu = 0$ and $\nu = 1$ regions a narrow compressible strip exists, following a constant energy contour of the bare disorder potential. **c**, A generalized phase diagram for QH-localization as deduced from our measurements. At low B , localized states are dominated by disorder and follow the single-particle phenomenology. Here the familiar QH phase transitions at half-filling of Landau levels are observed. At higher fields, interactions prevail and localized states form within dots or antidots. In this regime, new percolation phase boundaries emerge (red lines) and limit the QH phenomena to strips of width $\Delta n_{\text{disorder}}$ centred around their corresponding filling factors.

We summarize our results in a generalized phase diagram of localization in the QH regime (Fig. 4c). The disorder in any given sample is characterized by $\Delta n_{\text{disorder}}$ irrespective of the density. Depending on the interaction strength relative to the disorder, localized states acquire a different nature. When disorder is strong or when B is small ($\Delta n_{\text{disorder}} \gg n_{\text{max}}$), localized states are narrow compressible strips, which obey the single-particle phenomenology. Consequently, the familiar QH phase transitions exist at half-filling of the Landau levels. This explains why single-particle scaling of QH transitions is observed only in highly-disordered samples^{3–9}. At higher magnetic fields, localization is dominated by interactions and localized states form within dots or antidots. In this regime, new percolation phase boundaries emerge (red lines in Fig. 4c), limiting the QH phenomena to strips of width $\Delta n_{\text{disorder}}$ centred around the QH filling factors. □

Received 26 August; accepted 21 November 2003; doi:10.1038/nature02230.

1. Prange, R. E. & Girvin, S. M. *The Quantum Hall Effect* (Springer, New York, 1990).
2. Hückestein, B. Scaling theory of the integer quantum Hall-effect. *Rev. Mod. Phys.* **67**, 357–396 (1995).
3. Wei, H., Tsui, D., Paalanen, M. & Pruisken, A. Experiments on delocalization and universality in the integral quantum Hall effect. *Phys. Rev. Lett.* **61**, 1294–1296 (1988).

4. Wei, H. P., Lin, S. Y., Tsui, D. C. & Pruisken, A. M. M. Effect of long-range potential fluctuations on scaling in the integer quantum Hall-effect. *Phys. Rev. B* **45**, 3926–3928 (1992).
5. Engel, L., Wei, H. P., Tsui, D. C. & Shayeghan, M. Critical exponent in the fractional quantum Hall-effect. *Surf. Sci.* **229**, 13–15 (1990).
6. Koch, S., Haug, R., von-Klitzing, K. & Ploog, K. Size-dependent analysis of the metal-insulator transition in the integral quantum Hall effect. *Phys. Rev. Lett.* **67**, 883–886 (1991).
7. Engel, L. W., Shahar, D., Kurdak, C. & Tsui, D. C. Microwave frequency-dependence of integer quantum Hall-effect—evidence for finite-frequency scaling. *Phys. Rev. Lett.* **71**, 2638–2641 (1993).
8. Hohls, F., Zeitler, U. & Haug, R. Hopping conductivity in the quantum Hall effect: Revival of universal scaling. *Phys. Rev. Lett.* **88**, 1–4 (2002).
9. Hohls, F. *et al.* Dynamical scaling of the quantum Hall plateau transition. *Phys. Rev. Lett.* **89**, 1–4 (2002).
10. Shahar, D. *et al.* A new transport regime in the quantum Hall effect. *Solid State Commun.* **107**, 19–23 (1998).
11. Balaban, N., Meirav, U. & Bar-Joseph, I. Absence of scaling in the integer quantum Hall effect. *Phys. Rev. Lett.* **81**, 4967–4970 (1998).
12. Cobden, D. H., Barnes, C. H. W. & Ford, C. J. B. Fluctuations and evidence for charging in the quantum Hall effect. *Phys. Rev. Lett.* **82**, 4695–4698 (1999).
13. Chklovskii, D. B. & Lee, P. A. Transport properties between quantum Hall plateaus. *Phys. Rev. B* **48**, 18060–18078 (1993).
14. Cooper, N. R. & Chalker, J. T. Coulomb interactions and the integer quantum Hall-effect—screening and transport. *Phys. Rev. B* **48**, 4530–4544 (1993).
15. Ruzin, I., Cooper, N. & Halperin, B. Nonuniversal behavior of finite quantum Hall systems as a result of weak macroscopic inhomogeneities. *Phys. Rev. B* **53**, 1558–1572 (1996).
16. Ilani, S., Yacoby, A., Mahalu, D. & Shtrikman, H. Unexpected behavior of the local compressibility near the $B = 0$ metal-insulator transition. *Phys. Rev. Lett.* **84**, 3133–3136 (2000).
17. Yoo, M. J. *et al.* Scanning single-electron transistor microscopy: Imaging individual charges. *Science* **276**, 579–582 (1997).
18. Yacoby, A., Hess, H. F., Fulton, T. A., Pfeiffer, L. N. & West, K. W. Electrical imaging of the quantum Hall state. *Solid State Commun.* **111**, 1–13 (1999).
19. Zhitenev, N. B. *et al.* Imaging of localized electronic states in the quantum Hall regime. *Nature* **404**, 473–476 (2000).
20. Ilani, S., Yacoby, A., Mahalu, D. & Shtrikman, H. Microscopic structure of the metal-insulator transition in two dimensions. *Science* **292**, 1354–1357 (2001).
21. Efros, A. L. & Ioffe, A. F. Nonlinear screening and the background density of 2DEG states in magnetic field. *Solid State Commun.* **67**, 1019–1022 (1988).
22. Efros, A. L., Pikus, F. G. & Burnett, V. G. Density of states of a two-dimensional electron gas in a long-range random potential. *Phys. Rev. B* **47**, 2233–2243 (1993).
23. Shashkin, A. *et al.* Percolation metal-insulator transitions in the two-dimensional electron system of AlGaAs/GaAs heterostructures. *Phys. Rev. Lett.* **73**, 3141–3144 (1994).
24. Kukushkin, I., Fal'ko, V., Haug, R., von Klitzing, K. & Eberl, K. Magneto-optical evidence of the percolation nature of the metal-insulator transition in the two-dimensional electron system. *Phys. Rev. B* **53**, R13260–R13263 (1996).

Acknowledgements We benefited from discussions with A. M. Fink'elstein, Y. Gefen, Y. Meir, A. Stern and N. B. Zhitenev. This work was supported by the Israeli Science Foundation and the German MINERVA foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.I. (shahal.ilani@cornell.edu).

The role of increasing temperature variability in European summer heatwaves

Christoph Schär¹, Pier Luigi Vidale¹, Daniel Lüthi¹, Christoph Frei¹, Christian Häberli², Mark A. Liniger² & Christof Appenzeller²

¹Atmospheric and Climate Science, ETH Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

²MeteoSwiss, Krähbühlstrasse 58, 8044 Zürich, Switzerland

Instrumental observations^{1,2} and reconstructions^{3,4} of global and hemispheric temperature evolution reveal a pronounced warming during the past ~150 years. One expression of this warming is the observed increase in the occurrence of heatwaves^{5,6}. Conceptually this increase is understood as a shift of the statistical distribution towards warmer temperatures, while changes in the width of the distribution are often considered small⁷. Here we

show that this framework fails to explain the record-breaking central European summer temperatures in 2003, although it is consistent with observations from previous years. We find that an event like that of summer 2003 is statistically extremely unlikely, even when the observed warming is taken into account. We propose that a regime with an increased variability of temperatures (in addition to increases in mean temperature) may be able to account for summer 2003. To test this proposal, we simulate possible future European climate with a regional climate model in a scenario with increased atmospheric greenhouse-gas concentrations, and find that temperature variability increases by up to 100%, with maximum changes in central and eastern Europe.

A record-breaking heatwave affected the European continent in summer 2003. In a large area, mean summer (June, July and August, referred to as JJA below) temperatures have exceeded the 1961–90 mean by $\sim 3^\circ\text{C}$, corresponding to an excess of up to 5 standard deviations (Fig. 1a). Even away from the centre of action, many long-standing temperature records have tumbled.

For further analysis, we consider long-term temperature series

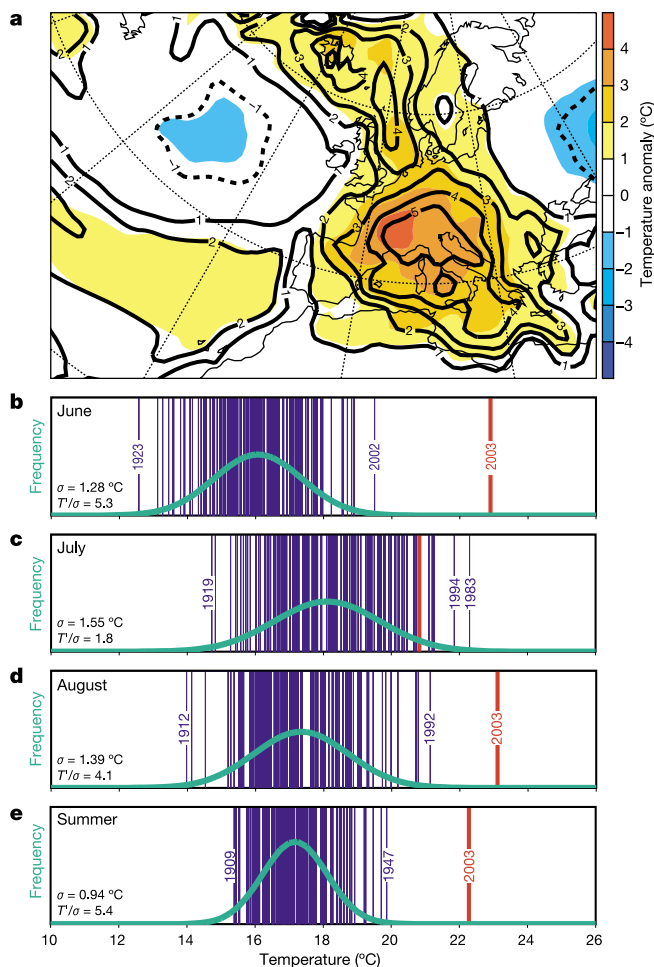


Figure 1 Characteristics of the summer 2003 heatwave. **a**, JJA temperature anomaly with respect to the 1961–90 mean. Colour shading shows temperature anomaly ($^\circ\text{C}$), bold contours display anomalies normalized by the 30-yr standard deviation. **b–e**, Distribution of Swiss monthly and seasonal summer temperatures for 1864–2003. The fitted gaussian distribution is indicated in green. The values in the lower left corner of each panel list the standard deviation (σ) and the 2003 anomaly normalized by the 1864–2000 standard deviation (T'/σ). See Methods section for further details.

from Switzerland, located close to the centre of the anomaly. Twelve carefully homogenized series^{8,9} are available with daily resolution since 1864. To minimize contamination by local meteorological and instrumental conditions, we amalgamate four independent and particularly reliable stations (Basel-Binningen, Geneva, Bern-Liebfeld, and Zürich) into one single series with monthly temporal resolution. This series is representative for the northwestern foothills of the Alps. Figure 1b–e displays the statistical distribution of monthly and seasonal temperatures. The year 2003 is far off the distribution in three of the four panels. For instance, the previous record holder for JJA was 1947 with a temperature anomaly of $T' = 2.7^\circ\text{C}$ (with respect to the 1864–2000 mean). The corresponding value for 2003 is as high as $T' = 5.1^\circ\text{C}$ and this amounts to an offset of 5.4 standard deviations from the mean (the corresponding values of individual months are listed in Fig. 1). Such extreme values (which indeed have the characteristics of outliers) pose serious challenges to any analysis, as the statistical distribution so far away from the mean is not described by the data.

In a first step, we thus restrict attention to the time period 1864–2000 and compile compound statistics for all monthly temperature anomalies (January–December). The purpose of this is to identify changes near the tails of the statistical distribution that result from the warming trend in the series. To this end, we consider two 60-yr periods, one covering the beginning of the series (1864–1923), and one the end (1941–2000). Figure 2a, b shows the resulting statistical distributions, both in terms of cumulative probability and probability density functions. The two distributions show similar characteristics in general, but the 1941–2000 distribution is shifted by the mean warming ($\Delta T = 0.8^\circ\text{C}$) between the two periods. This shift also implies a change in the frequency of extremes. For instance (Fig. 2c), the frequency of a month with an anomaly of $T' = 3^\circ\text{C}$ has increased by $\sim 100\%$. Hence, a month in the 1941–2000 period with an excess temperature of $T' = 3^\circ\text{C}$ can be tied with a probability of 50% to the warming between the two periods, in a probabilistic sense as recently proposed¹⁰. This illustrates how comparatively small shifts in climate mean may imply pronounced changes at the tails of the statistical distribution and in the frequency of extremes.

The dashed and full curves in Fig. 2 relate to the empirical and the fitted gaussian distributions, respectively, and their close agreement shows that the gaussian distribution is an excellent approximation to the data. The small reduction in variability (Fig. 2b) is not statistically significant, and is entirely due to changes in the month of December (where the variability was substantially reduced owing to the absence of cold northeasterly weather types).

A conclusive analysis such as that in Fig. 2 is not feasible for summer 2003, as there is only one data point so far off the mean. To quantitatively assess the situation, we have estimated its return period. The return period τ is an estimate of the frequency of a particular event (or its exceedance) based on a stochastic concept. Here we employ a gaussian distribution fitted to JJA temperatures to estimate τ with respect to a selected reference period (see Methods section for details). With respect to the reference period 1864–2000, a return period of several million years is obtained, but such an excessive estimate based on a short series is dubious. To account for the warming in the last decades, we use a more recent reference period 1990–2002 (with $\Delta T = 1.25^\circ\text{C}$ warmer mean temperature, but assuming an unchanged standard deviation). With respect to this climatology, the resulting return period for summer 2003 still amounts to $\tau = 46,000$ yr. The uncertainty of this estimate is considerable, however, and the lower bound of the 90% confidence interval is $\tau = 9,000$ yr.

This large return period should not be overstated, and is here merely used to express the rareness of such an extreme summer with respect to the long-term instrumental series available. In particular, the analysis does not exclude the possibility that such warm summers might have occurred in the more distant historical past,

for instance in the Medieval Warm Period¹¹, in 1540^{12,13} or in 1757. It suggests, however, that an event like summer 2003 does not fit into the gaussian statistics spanned by the observations of the reference period, but might rather be associated with a transient change of the statistical distribution. This interpretation is consistent with the idea that small changes of the statistical distribution can yield pronounced changes in the incidence of extremes^{7,14}.

As a shift of the statistical distribution by the observed mean warming is unable to explain the record-breaking summer 2003, we hypothesize that the heatwave might be due to a change of the distribution's width, representing an increase in year-to-year variability. Support for this hypothesis comes from a regional climate model (RCM) driven by a greenhouse-gas scenario representing 2071–2100 conditions (SCEN). The scenario integration is compared against a control integration covering the period 1961–90 (CTRL). At the lateral boundaries, the RCM is driven by a model chain consisting of two general circulation models (GCMs; see Methods section for details). The use of a high-resolution RCM increases our ability to compare the results against observations.

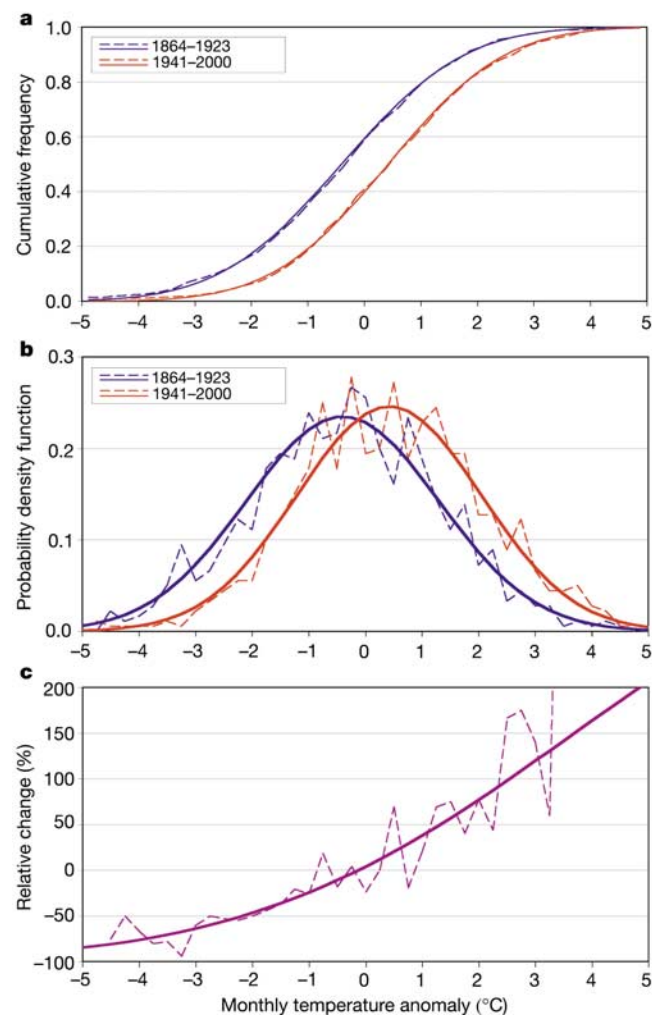


Figure 2 Statistical distribution of Swiss monthly temperature anomalies (compound statistics using January–December monthly data). Data in **a** and **b** are shown for the periods 1864–1923 (blue curves) and 1941–2000 (red curves). Panels show the cumulative frequency distribution (**a**), the probability density function (**b**), and the relative frequency change between the two periods (**c**). Full lines show the fit with the gaussian distribution, dashed lines are obtained from raw data.

The statistical temperature distribution for CTRL agrees notably well with observations. For the grid point in northern Switzerland, the summer climate is characterized by a temperature mean of $\bar{T} = 16.1^\circ\text{C}$ and a standard deviation of $\sigma = 0.96^\circ\text{C}$, while the long-term characteristics of our temperature series are $\bar{T} = 16.9^\circ\text{C}$ and $\sigma = 0.94^\circ\text{C}$.

Figure 3a, b displays JJA temperatures for the two integrations for a grid point in northern Switzerland. In the SCEN simulation, the distribution is shifted by $\sim 4.6^\circ\text{C}$ towards warmer temperatures. More important, SCEN also exhibits a pronounced widening of its statistical distribution, with the standard deviation increasing by 102%. This widening is statistically highly significant ($P < 1\%$) and only slightly affected by the transient warming within the two periods (a revised estimate using detrended temperature series yields a somewhat smaller variability increase of 86%). The spatial distribution of the relative increase in variability (Fig. 3d) shows a pronounced signal throughout central and eastern Europe that is not directly linked to the simulated mean temperature change (Fig. 3c). More detailed analysis suggests that the warm summers of SCEN show signs of drought¹⁵, with the semi-arid Mediterranean climate progressing towards central Europe. In SCEN, central Europe is more often (but not always) affected by summer droughts than in CTRL, and this implies an increase in variability. The drought conditions develop in response to large-scale anticyclonic forcing, and they nonlinearly amplify local temperature anomalies. During droughts the net balance of solar and infrared radiation is almost entirely balanced by local heating, while evapotranspiration

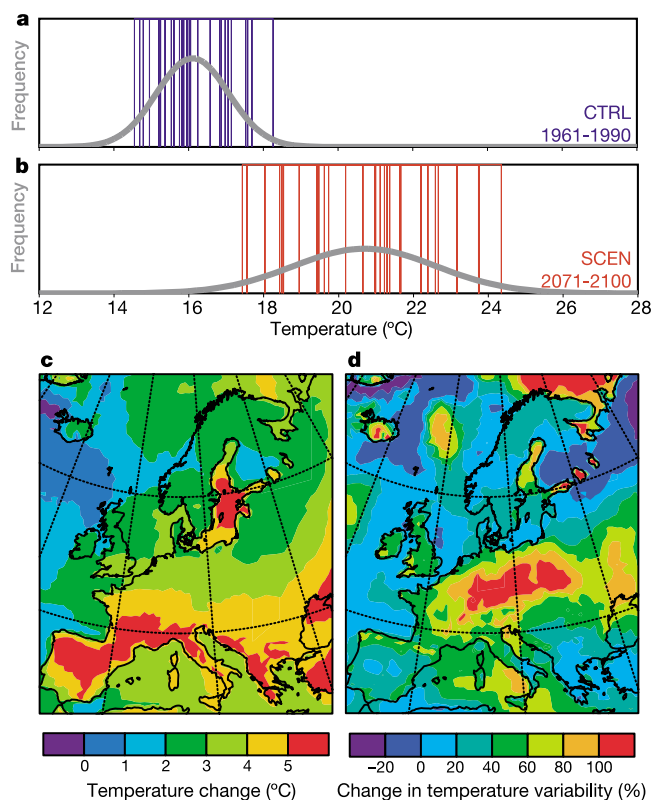


Figure 3 Results from an RCM climate change scenario representing current (CTRL 1961–90) and future (SCEN 2071–2100) conditions. **a**, **b**, Statistical distribution of summer temperatures at a grid point in northern Switzerland for CTRL and SCEN, respectively. **c**, Associated temperature change (SCEN–CTRL, $^\circ\text{C}$). **d**, Change in variability expressed as relative change in standard deviation of JJA means ((SCEN–CTRL)/CTRL, %).

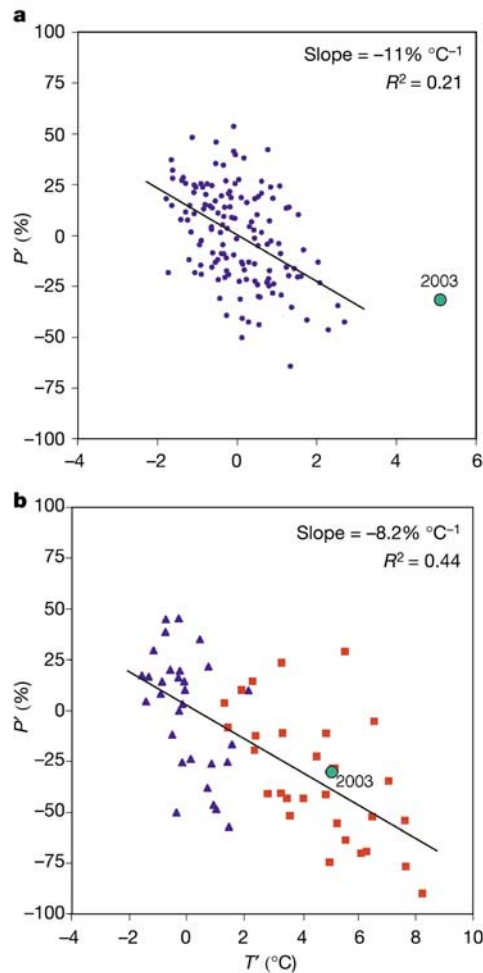


Figure 4 Scatter diagrams showing summer mean temperature and precipitation anomalies for northern Switzerland. **a**, Long-term (1864–2003) station data with respect to 1961–90 means. **b**, Climate change simulations CTRL (1961–90, blue symbols) and SCEN (2071–90, red symbols) with respect to CTRL means. The green symbols show the observations for JJA 2003. The regression lines in **a** and **b** are based on 1864–2002 data and combined CTRL and SCEN data, respectively.

is suppressed owing to the lack of soil moisture¹⁶. This process may be further amplified by a positive feedback between soil moisture and precipitation^{17,18}.

The sequence of feedbacks involves substantial uncertainties due to large-scale anticyclonic forcing¹⁹, radiation²⁰, soil hydrology²¹ and other processes, which are difficult to represent in climate models. To check on our simulations, we have analysed other GCM and RCM scenarios of greenhouse-gas conditions, and find that all of these exhibit a substantially increased level of variability over large parts of Europe (we have studied one GCM and four RCM simulations from the PRUDENCE project; <http://prudence.dmi.dk>).

The simulated increase in variability also implies an increase in extremes relative to mean climatic conditions. For illustration, a 50% increase in the standard deviation of our long-term JJA temperature series ($\sigma = 0.94^\circ\text{C}$) would raise the probability of a 2003-like event ($T' = 3.85^\circ\text{C}$ with respect to 1990–2002) by a factor of ~ 150 . For an event with $T' = 5^\circ\text{C}$, it would increase by a factor of $\sim 5,100$. This tremendous sensitivity of extremes to the width of the statistical distribution has led to the statement “variability is more important than averages”²⁴. A recent increase in variability is thus a plausible hypothesis to explain extreme JJA 2003 conditions. Such a hypothesis would also be compatible with

the occurrence of drastically different European summers such as in 2002 and 2003, but at present there are insufficient data to draw any firm conclusions.

To conclude, we address the question of whether the abnormal summer 2003 shows similar characteristics to those simulated in the RCM runs. To this end, summer temperature and precipitation anomalies are displayed against each other, both for the observations (Fig. 4a) and for the climate change simulations (Fig. 4b). Both panels include a data point representing observed JJA 2003 conditions, and the results apply to northern Switzerland. The observed data (Fig. 4a) are based on averages of conventional temperature and precipitation (rain-gauge) observations at the four stations referred to above, while the simulated data (Fig. 4b) are shown for a single grid point roughly corresponding to the location of our long-term series.

Several inferences can be drawn from the analysis. First, both data sets exhibit a similar (statistically significant) relationship between temperature and precipitation anomalies. The regression analysis yields slopes of -11°C^{-1} and -8.2°C^{-1} for the observations and the simulations, respectively. Thus, although there is some underestimation of summer precipitation in CTRL (at the grid point under consideration, by 21%), the simulations credibly represent the observed precipitation sensitivity. Despite a general trend towards drier conditions with increasing temperatures, there is also an increase in the incidence of heavy precipitation events²².

Second, Fig. 4b demonstrates that in terms of temperature and precipitation the climatic conditions in JJA 2003 were not unlike those simulated by SCEN for the period 2071–2100. For northern Switzerland, the 2003 observation is located approximately in the middle of the SCEN data points (Fig. 4b). Thus, the RCM simulations suggest that towards the end of the century—under the given scenario assumptions—about every second summer could be as warm or warmer (and as dry or dryer) than 2003.

Our results demonstrate that the European summer climate might experience a pronounced increase in year-to-year variability in response to greenhouse-gas forcing. Such an increase in variability might be able to explain the unusual European summer 2003, and would strongly affect the incidence of heatwaves and droughts in the future. It would represent a serious challenge to adaptive response strategies designed to cope with climate change. □

Methods

Large-scale analysis of summer 2003

The continental-scale temperature anomaly for JJA 2003 (Fig. 1a) is based on ERA-40 reanalysis data²³ (for 1961–90) and operational meteorological analysis data (for 2003) of the European Centre for Medium-Range Weather Forecasts (ECMWF; see <http://www.ecmwf.int>). Monthly temperatures are computed as means of daily $T_{\min}$ and $T_{\max}$. Small height differences between the ERA-40 and ECMWF topographies are accounted for by the use of an adiabatic lapse rate (0.6°C per 100 m).

Estimation of return period

The stochastic concept adopted in the estimation of return periods assumes independent, identically distributed JJA temperatures with the underlying distribution being gaussian. The distribution parameters are estimated from the data of the reference period, using the method of moments (which in the case of a gaussian distribution is identical to maximum-likelihood estimation). The return period of the event (expected frequency of threshold exceedance) is then calculated from the fitted distribution. Confidence bounds of the return period were calculated by parametric resampling. These take into account the uncertainty of the parameter estimates given the finite sample size (that is, the number of summers in the reference period), but not the uncertainty in the underlying stochastic concept. We have also tested whether the data are reasonably gaussian distributed, checking quantile-quantile plots (see also Fig. 2a).

Climate change simulations

The climate change scenario is based on the SRES A2 transient greenhouse-gas scenario as specified by the Intergovernmental Panel on Climate Change (IPCC)²⁴. The scenario computations involve three numerical models: the low-resolution HadCM3 global coupled atmosphere–ocean GCM, the intermediate-resolution HadAM3H atmospheric GCM, and the CHRM limited-area high-resolution RCM. The HadCM3 simulation^{25,26} is a long integration using the observed atmospheric composition for 1859–1990 and scenario conditions for 1991–2100. For the HadAM3H simulation²⁷, two time-slice

experiments are available, representing control (1961–90) and scenario (2071–2100) conditions. The former is driven by observed sea surface temperature and sea-ice distributions, while the latter uses the changes from the HadCM3²⁸. The CHRM RCM²⁹ is used with a horizontal resolution of 56 km and 20 levels in the vertical, and is driven at its lateral boundaries by HadAM3H. The CHRM has been validated regarding its ability to represent observed natural interannual variations²⁹. The simulated increase in temperature variability in SCEN is largely determined by the soil hydrology of the model under consideration. For instance, it is substantially smaller in the CHRM than in the driving HadAM3H simulation.

Received 15 September; accepted 17 December 2003; doi:10.1038/nature02300.

Published online 11 January 2004.

- Jones, P. D. & Moberg, A. Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Clim.* **16**, 206–223 (2003).
- Folland, C. K. *et al.* Global temperature change and its uncertainties since 1861. *Geophys. Res. Lett.* **28**, 2621–2624 (2001).
- Crowley, T. J. Causes of climate change over the past 1000 years. *Science* **289**, 270–277 (2000).
- Jones, P. D., Osborn, T. J. & Briffa, K. R. The evolution of climate over the last millennium. *Science* **292**, 662–667 (2001).
- Easterling, D. R. *et al.* Observed variability and trends in extreme climate events: A brief review. *Bull. Am. Meteorol. Soc.* **81**, 417–425 (2000).
- Frich, P. *et al.* Observed coherent changes in climatic extremes during the second half of the twentieth century. *Clim. Res.* **19**, 193–212 (2002).
- Folland, C. K. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 99–182 (Cambridge Univ. Press, Cambridge, UK, 2001).
- Aschwanden, A. *et al.* *Bereinigte Zeitreihen: Die Ergebnisse des Projekts KLIMA90* (MeteoSwiss, Zürich, 1996).
- Bergert, M. *et al.* *Homogenisierung von Klimamessreihen und Berechnung der Normwerte 1961–1990* Veröffentlichungen der MeteoSchweiz, 67, MeteoSwiss, Zürich, (2003).
- Allen, M. R. Liability for climate change. *Nature* **42**, 891–892 (2003).
- Esper, J., Cook, E. R. & Schweingruber, F. H. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science* **295**, 2250–2253 (2002).
- Pfister, C. *et al.* Documentary evidence on climate in sixteenth-century Europe. *Clim. Change* **43**, 55–110 (1999).
- Jacobeit, J., Wanner, H., Koslowski, G. & Gudd, M. European surface pressure patterns for months with outstanding climatic anomalies during the sixteenth century. *Clim. Change* **43**, 201–221 (1999).
- Katz, R. W. & Brown, B. G. Extreme events in a changing climate: Variability is more important than averages. *Clim. Change* **21**, 289–302 (1992).
- Wetherald, R. T. & Manabe, S. The mechanisms of summer dryness induced by greenhouse warming. *J. Clim.* **8**, 3096–3108 (1995).
- Hartmann, D. L. *Global Physical Climatology* (Academic, San Diego, 1994).
- Enthekabi, D., Rodriguez-Iturbe, I. & Bras, R. L. Variability in large-scale water balance with land surface-atmosphere interaction. *J. Clim.* **5**, 798–813 (1992).
- Schär, C., Lüthi, D., Beyerle, U. & Heise, E. The soil-precipitation feedback: A process study with a regional climate model. *J. Clim.* **12**, 722–741 (1999).
- Rodwell, M. J. & Hoskins, B. J. Subtropical anticyclones and summer monsoons. *J. Clim.* **14**, 3192–3211 (2001).
- Wild, M., Dümenil, L. & Schulz, J. P. Regional climate simulation with a high resolution GCM: surface hydrology. *Clim. Dyn.* **12**, 755–774 (1996).
- Seneviratne, S. I., Pal, J. S., Eltahir, E. A. B. & Schär, C. Summer dryness in the US midwest: A process study with a regional climate model. *Clim. Dyn.* **20**, 69–85 (2002).
- Christensen, J. H. & Christensen, O. B. Severe summertime flooding in Europe. *Nature* **421**, 805–806 (2002).
- Simmons, A. J. & Gibson, J. K. *The ERA-40 Project Plan* ERA-40 (Project Report Series, No. 1, ECMWF, Reading, 2000).
- Nakicenovic, N. *et al.* *IPCC Special Report on Emissions Scenarios* (Cambridge Univ. Press, Cambridge, UK, 2000).
- Johns, T. C. *et al.* Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Clim. Dyn.* **20**, 583–612 (2003).
- Collins, M., Tett, S. F. B. & Cooper, C. The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **17**, 61–81 (2001).
- Pope, D. V., Gallani, M., Rowntree, R. & Stratton, A. The impact of new physical parameterizations in the Hadley Centre climate model HadAM3. *Clim. Dyn.* **16**, 123–146 (2000).
- Jones, R. G., Murphy, J. M., Hassell, D. C. & Taylor, R. *Ensemble Mean Changes in a Simulation of the European Climate of 2071–2100 Using the New Hadley Centre Regional Modelling system HadAM3H/HadRM3H* (Hadley Centre Report, Hadley Centre, Exeter, UK, 2001); available at (<http://prudence.dmi.dk>) (2001).
- Vidale, P. L., Lüthi, D., Frei, C., Seneviratne, S. & Schär, C. Predictability and uncertainty in a regional climate model. *J. Geophys. Res.* **D 108** doi:10.1029/2002JD002810 (2003).

Acknowledgements We thank the climatology and data teams of MeteoSwiss (Zürich, Switzerland) for providing access to their long-term homogenized temperature series, the Hadley Centre (Exeter, UK) for providing access to their climate change simulations, and the PRUDENCE team for access to climate simulations. We also thank our colleagues for comments on the manuscript. The computations were performed on the computing facilities of ETH and the Swiss Center for Scientific Computing (CSCS). This research was supported by the Fifth Framework Programme of the European Union (project PRUDENCE), by the Swiss Ministry for Education and Research, and by the Swiss National Science Foundation (NCCR Climate).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.S. (schaer@env.ethz.ch).

Reburial of fossil organic carbon in marine sediments

Angela F. Dickens^{1,2}, Yves Gélinas³, Caroline A. Masiello^{4,5}, Stuart Wakeham⁶ & John I. Hedges¹

¹School of Oceanography, University of Washington, Box 355351, Seattle, Washington 98195-5351, USA

²Department of Chemistry, University of Washington, Box 351700, Seattle, Washington 98195-1700, USA

³Chemistry and Biochemistry Department, Concordia University, 1455 de Maisonneuve Boulevard West, Montreal (Quebec), Canada H3G 1M8

⁴Division of Geology and Planetary Sciences, Mail Code 170-25, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA

⁵Department of Geography, University of California Santa Barbara, Santa Barbara, California 93106, USA

⁶Skidaway Institute of Oceanography, 10 Ocean Science Circle, Savannah, Georgia 31411, USA

Marine sediments act as the ultimate sink for organic carbon, sequestering otherwise rapidly cycling carbon for geologic time-scales^{1,2}. Sedimentary organic carbon burial appears to be controlled by oxygen exposure time *in situ*^{3,4}, and much research has focused on understanding the mechanisms of preservation of organic carbon⁵. In this context, combustion-derived black carbon has received attention as a form of refractory organic carbon that may be preferentially preserved in soils^{6,7} and sediments^{8,9}. However, little is understood about the environmental roles, transport and distribution of black carbon. Here we apply isotopic analyses to graphitic black carbon samples isolated from pre-industrial marine and terrestrial sediments. We find that this material is terrestrially derived and almost entirely depleted of radiocarbon, suggesting that it is graphite weathered from rocks, rather than a combustion product. The widespread presence of fossil graphitic black carbon in sediments has therefore probably led to significant overestimates of burial of combustion-derived black carbon in marine sediments. It could be responsible for biasing radiocarbon dating of sedimentary organic carbon, and also reveals a closed loop in the carbon cycle. Depending on its susceptibility to oxidation, this recycled carbon may be locked away from the biologically mediated carbon cycle for many geologic cycles.

Black carbon (BC) is the heterogeneous, aromatic and carbon-rich residue of biomass burning and fossil-fuel combustion^{10,11}. It is formed on land, and may be eroded from soils and carried by rivers or via aerosol transport to the ocean. BC in open-ocean sediments is from 2,400 to 13,900 yr older than the co-deposited bulk organic carbon (OC)¹². It is possible that BC is small and light enough to be stored in the oceanic dissolved organic carbon pool before being deposited in marine sediments, accounting for the older age of the BC. Alternatively, BC may age on land before transport to the ocean, either in the soil carbon pool or as fossil BC from ancient forest fires stored in rocks. Existing data cannot distinguish between these storage and transport mechanisms.

We examined marine sediments collected along a well-characterized transect perpendicular to the Washington coast⁴, ranging from terrestrially influenced nearshore sites to marine-dominated offshore sediments. We determined sedimentary mixed layer depth using ²¹⁰Pb and sampled horizons (9–23 cm) deep enough beneath this layer to be pre-industrial according to ¹⁴C sedimentation rates. The low levels of nuclides from nuclear fallout¹³ and background levels of combustion-derived polycyclic aromatic hydrocarbons¹⁴ present at our sampling depths in similar Washington slope samples further suggest that deeply mixed anthropogenic graphitic black carbon (GBC) is minimal in our samples. We extracted GBC, an

extremely refractory form of BC, from sediments by stepwise demineralization, hydrolysis of reactive organic matter and thermal oxidation⁸. This method isolates only very highly condensed soot-BC and graphite but destroys charcoal-BC and kerogen (non-graphitized rock carbon).

GBC concentrations increase roughly linearly with increasing distance offshore, when plotted both as weight per cent of the sediment (Fig. 1a) and as per cent of total organic carbon (TOC, Fig. 1a inset), reaching 6.5% of TOC. However, the flux of GBC to sediments decreases with increasing distance offshore (Fig. 1b), suggesting that the GBC is terrestrially derived and that the concentration trends arise because GBC is preferentially transported offshore relative to sedimentary minerals and TOC.

We analysed the bulk OC and GBC fractions of a subset of the Washington coast sediments for ^{13}C and ^{14}C signatures. Radiocarbon was measured via accelerator mass spectrometry¹⁵ and is reported as an age-corrected Δ value¹⁶, corresponding to the ^{14}C signature at the time of deposition. Pre-industrial modern carbon has a Δ value of 0‰ and fossil material has a value of about -1,000‰. For bulk OC, $\delta^{13}\text{C}$ signatures range from -24.3‰ nearshore to -21.1‰ offshore, and Δ values range from -9‰ nearshore to -240‰ offshore, corresponding to ^{14}C ages of between 30 and 2,220 yr (Table 1). These isotopic data suggest that nearshore sediments are dominated by modern carbon, with a significant amount of terrestrial C_3 plant material ($\delta^{13}\text{C} \approx -27$ ‰), whereas offshore the sediments are older and contain mostly marine material ($\delta^{13}\text{C} \approx -21$ ‰).

For GBC, the Δ values range between -895‰ and -989‰, approaching the fossil carbon limit, whereas the $\delta^{13}\text{C}$ values of -19.4‰ to -21.3‰ fall within the range for marine plankton (Table 1). These values are uniform with distance offshore, and correspond to ^{14}C ages of between 18,600 and 37,450 yr (Table 1). The existence of a biospheric reservoir in which an entire pool of carbon could age for more than 35,000 yr, undiluted by younger

carbon, is extremely unlikely, indicating that the GBC fraction consists predominantly of fossil carbon. The Washington coast GBC samples have Δ and $\delta^{13}\text{C}$ values that cluster together (Fig. 2), implying uniformity of the source material. Whereas GBC has a similar ^{13}C signature to marine plankton, its radiocarbon signature dramatically separates it from both contemporary marine and terrestrial sources, making these unlikely endmembers for this material.

We also isolated GBC from three terrestrial sediments: two pre-industrial horizons from Lake Washington¹⁷ and surface sediments from the Stillaguamish River, in the northwestern Washington State, USA, and from an open ocean site in the Equatorial Pacific at 9° N (EqPac)¹⁸. The GBC fractions from the terrestrial sites were slightly depleted of ^{13}C compared with the Washington coast samples but had virtually identical ^{14}C signatures (Fig. 2). The similarity of GBC isotopic signatures from both terrestrial and marine sediments suggests a common, terrestrial GBC source. GBC from the EqPac plots near but outside the GBC cluster, with a slightly more enriched $\delta^{13}\text{C}$ signature and more positive Δ value (Fig. 2). The similarity of isotopic signatures suggests that the EqPac GBC contains a fraction of the same radiocarbon-dead material. However, this GBC sample must also contain a substantial component of a younger, ^{13}C -enriched material to account for the differences in isotopic signature compared with the Washington coast sediments. One possible source of such material is soot from combustion of C_4 grasses¹⁹. The ^{14}C -depletion previously observed in Southern Ocean BC¹² suggests that radiocarbon-dead GBC is present in this site as well and, in combination with our results, implies that radiocarbon-dead GBC is important in sediments world-wide.

The measured isotopic signatures of the GBC fraction of sediments appear robust and widespread, indicating that the GBC is a terrestrially derived material that is fossil in age and has a marine-like ^{13}C signature. Because all sediments (except the river sediment) were sampled from pre-industrial horizons, fossil-fuel-derived soot was not a major GBC source. A possible GBC source that satisfies all of the evidence is petrogenic graphite. This material would be weathered from rocks on land, would have a fossil Δ value, and could have a marine $\delta^{13}\text{C}$ signature, depending on its ultimate sources and the conditions during lithification. Graphite dispersed in metamorphic rocks tends to have a biogenic ^{13}C signature (between -18‰ and -30‰) but may be as ^{13}C -enriched as 0‰ (refs 20, 21). The observed range of GBC ^{13}C signatures could reflect

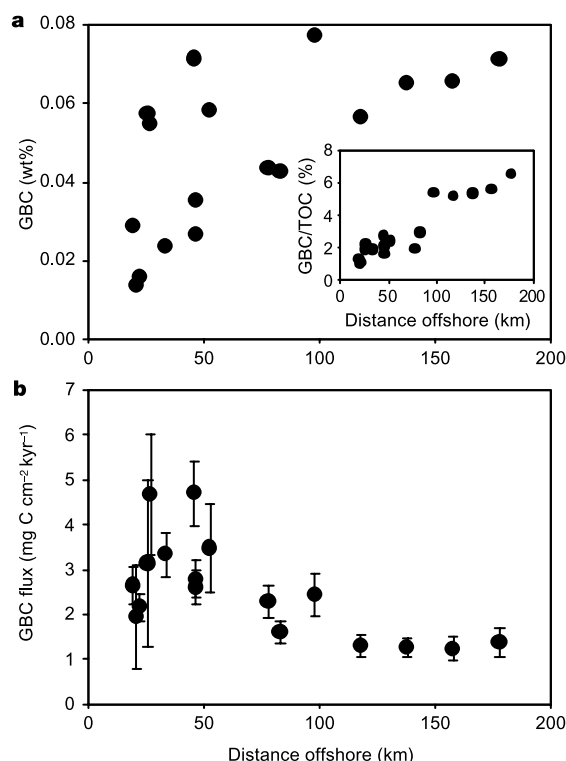


Figure 1 Concentrations and fluxes of GBC off the Washington coast. **a**, GBC wt% concentrations versus distance offshore. The inset shows the same data plotted as per cent of total organic carbon (TOC). **b**, Fluxes of GBC as a function of distance offshore.

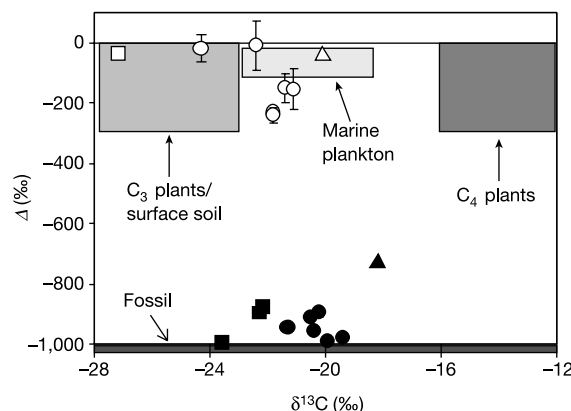


Figure 2 Plot of Δ versus $\delta^{13}\text{C}$ for all TOC (open symbols) and GBC (closed symbols) samples. Washington coast samples are shown as circles, the three terrestrial samples as squares and the one open-ocean sample as a triangle. Error bars for the Washington coast samples are smaller than the symbol if not visible. Error bars for the other samples are not shown. Approximate ranges of isotopic signatures for possible endmembers are indicated as grey rectangles, and are based on a review of recent literature.

Table 1 Concentrations, fluxes and isotopic signatures of TOC and GBC in sediments

Sample	Distance offshore (km)	Sedim. rate* (cm kyr ⁻¹)	Fraction	wt%	δ ¹³ C (‰)	Δ (‰)	¹⁴ C age (yr)	GBC flux (mg cm ⁻² kyr ⁻¹)
WA coast 4	20.8	16.2 ± 9.4	TOC	1.38	-24.3	-19 ± 44	110 ± 370	
			GBC	0.014	-20.5	-910 ± 9	19,840 ± 860	1.9 ± 1.2
WA coast 15	25.6	15.9 ± 9.1	TOC	2.99	-22.4	-9 ± 84	30 ± 700	
			GBC	0.057	-20.2	-895 ± 14	18,600 ± 1,100	3.1 ± 1.8
WA coast 6	46.5	13.5 ± 0.2	TOC	1.66	-21.8	-232 ± 3	2,130 ± 30	
			GBC	0.035	-19.4	-979 ± 5	31,900 ± 2,000	2.8 ± 0.4
WA coast 9	117.7	5.6 ± 0.7	TOC	1.10	-21.8	-240 ± 26	2,220 ± 290	
			GBC	0.057	-19.9	-989 ± 1	37,450 ± 670	1.3 ± 0.2
WA coast 1	157.8	4.7 ± 0.7	TOC	1.17	-21.4	-149 ± 47	1,280 ± 450	
			GBC	0.065	-21.3	-942 ± 3	23,510 ± 470	1.2 ± 0.3
WA coast 16	177.8	4.7 ± 0.9	TOC	1.09	-21.1	-155 ± 67	1,340 ± 650	
			GBC	0.071	-20.4	-957 ± 4	25,880 ± 680	1.4 ± 0.3
EqPac, 9°N		1	TOC	0.35	-20.1	-41	298	
			GBC	0.0092	-18.2	-731	10,820	0.048
LW 50–62		100	TOC	3.26	-27.2	n.d.	n.d.	
			GBC	0.031	-22.3	-893	18,404	15.5
LW 95–300		100	TOC	4.62	-27.2	-35	241	
			GBC	0.021	-22.2	-874	17,104	10.5
Still. River		n.a.	TOC	0.142	n.d.	n.d.	n.d.	
			GBC	0.044	-23.6	-994.1 ± 0.1	41,260 ± 160	n.d.

Δ is the age-corrected Δ¹⁴C value corresponding to the ¹⁴C signature at the time of deposition¹⁶. Standard deviations in GBC concentrations were conservatively estimated to be ±15%. Standard deviation for all δ¹³C measurements is ±0.1‰. GBC fluxes were calculated according to the formula: flux = (wt% GBC) × sedimentation rate × bulk density. Standard deviations for Δ and ¹⁴C age reflect uncertainty associated with statistics of the accelerator mass spectrometry (AMS) measurement and age-correction calculations. ¹⁴C age was calculated using the real mean life for ¹⁴C (8,267 yr). Sedim. rate, sedimentation rate; wt%, weight per cent; WA, Washington; EqPac, Equatorial Pacific; LW, Lake Washington; Still. River, Stillaguamish River; n.d., not determined; n.a., not applicable.

*Sedimentation rates in bold were calculated from downcore profiles of ¹⁴C reported in ref. 4, except for WA coast no. 4 which was measured for this study. Other WA coast sedimentation rates were calculated by assuming a linear decrease in sedimentation rate with distance offshore and extrapolating from the measured rates. The sedimentation rate for the EqPac site is an estimate based on our TOC radiocarbon age. This value is an order of magnitude larger than that estimated in ref. 29, probably owing to deeply mixed radiocarbon³⁰. The rate for the LW site is from ref. 17.

weathering of different sources of graphite on land. Kerogen is an unlikely GBC source because even highly aromatic type II kerogens²² yielded less than 0.6% of their carbon as GBC (Supplementary Table S1).

Previous reports of BC concentrations and fluxes in ocean sediments^{23,24} may have included petrogenic graphite in addition to combustion-derived BC, leading to significant overestimates of BC derived from biomass burning. Such overestimates could compromise measurements and interpretations of budgets for BC burial in marine sediments. For example, our estimated GBC flux for the EqPac site is 30–80% of the BC fluxes reported²⁴ for comparable distances from the continents (~3,000 km). On the basis of these numbers and estimating that the EqPac GBC is 70% fossil, it appears that a significant fraction (~20–60%) of these published open-ocean BC fluxes might consist of petrogenic GBC. Overestimation of BC burial implies an underestimation of BC degradation rates in the environment, and further complicates our understanding of BC dynamics and its role in the carbon cycle.

Our results imply that a significant fraction of sedimentary carbon is radiocarbon-dead GBC. This ¹⁴C-free component will significantly affect ¹⁴C measurements of bulk OC, making measured ¹⁴C ages for OC appear substantially older than their 'true' age. Using a simple isotopic mixing equation, we estimate that the presence of fossil GBC in our Washington coast sediments makes the TOC appear to be 75–530 yr older than the GBC-free OC would be. This bias in the ¹⁴C-derived TOC ages leads to overestimates of OC storage in intermediate reservoirs such as soils before deposition in marine sediments, and makes it more difficult to understand the timescales of OC transport between carbon reservoirs.

Finally, it appears that a fraction of petrogenic GBC is being weathered from rocks, carried to the ocean and redeposited in sediments. Several other studies have found indirect evidence of such reburial of fossil carbon or suggested that it might occur^{25–27}, but to the best of our knowledge this is the first study to have isolated an almost entirely radiocarbon-dead fraction from pre-industrial modern sediments. Petsch *et al.*²⁸ found that the OC in carbon-rich black shales was mostly but not entirely degraded in the environment, and suggested that a combination of slow degradation kinetics and fast erosion rates may be responsible for preservation of some fossil carbon. Similar mechanisms probably account for the presence of petrogenic GBC in sediments. This reburial of relatively

unaltered fossil carbon amounts to a closed loop in the Earth's carbon cycle.

Using an open-ocean GBC flux of 0.05 mg C cm⁻² kyr⁻¹ (estimated from the EqPac site) and a flux of 1.5 mg C cm⁻² kyr⁻¹ on the continental margins (from the offshore Washington coast sites), we estimate that a total of 7 × 10¹¹ g of fossil GBC is deposited yearly in the world ocean. This flux accounts for approximately 0.5% of the total burial of OC in marine sediments². If GBC is relatively unsusceptible to degradation, as seems likely, only a small fraction of GBC would be degraded during each geologic cycle. Assuming steady state, this implies that only a small amount of fossil GBC would be newly formed through metamorphism (that is, enough to balance GBC degradation). Slow rates of formation and destruction of GBC imply that a relatively large amount of GBC would be continually cycling in this GBC loop, which appears to act as a long-term OC holding pool away from the main streams of the carbon cycle. □

Received 25 September; accepted 16 December 2003; doi:10.1038/nature02299.

- Hedges, J. I. & Keil, R. G. Sedimentary organic matter preservation: an assessment and speculative synthesis. *Mar. Chem.* **49**, 81–115 (1995).
- Berner, R. A. Biogeochemical cycles of carbon and sulfur and their effect on atmospheric oxygen over Phanerozoic time. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **75**, 97–122 (1989).
- Hartnett, H. E., Keil, R. G., Hedges, J. I. & Devol, A. H. Influence of oxygen exposure time on organic carbon preservation in continental margin sediments. *Nature* **391**, 572–574 (1998).
- Hedges, J. I. *et al.* Sedimentary organic matter preservation: a test for selective degradation underoxic conditions. *Am. J. Sci.* **299**, 529–555 (1999).
- Hedges, J. I. *et al.* The molecularly-uncharacterized component of nonliving organic matter in natural environments. *Org. Geochem.* **31**, 945–958 (2000).
- Skjemstad, J. O., Taylor, J. A. & Smernik, R. J. Estimation of charcoal (char) in soils. *Commun. Soil Sci. Plant Anal.* **30**, 2283–2298 (1999).
- Schmidt, M. W. I., Skjemstad, J. O., Gehrt, E. & Kögel-Knabner, I. Charred organic carbon in German chernozemic soils. *Eur. J. Soil Sci.* **50**, 351–365 (1999).
- Gélinas, Y., Prentice, K. M., Baldock, J. A. & Hedges, J. I. An improved thermal oxidation method for the quantification of soot/graphitic black carbon in sediments and soils. *Environ. Sci. Technol.* **35**, 3519–3525 (2001).
- Middleburg, J. J., Nieuwenhuize, J. & van Breugel, P. Black carbon in marine sediments. *Mar. Chem.* **65**, 245–252 (1999).
- Goldberg, E. D. *Black Carbon in the Environment: Properties and Distribution* (Wiley & Sons, New York, 1985).
- Schmidt, M. W. I. & Noack, A. G. Black carbon in soils and sediments: analysis, distribution, implications, and current challenges. *Glob. Biogeochem. Cycles* **14**, 777–793 (2000).
- Masiello, C. A. & Druffel, E. R. M. Black carbon in deep-sea sediments. *Science* **280**, 1911–1913 (1998).
- Carpenter, R., Beasley, T. M., Zahnle, D. & Somayajulu, B. L. K. Cycling of fallout (Pu, ²⁴¹Am, ¹³⁷Cs) and natural (U, Th, ²¹⁰Pb) radionuclides in Washington continental slope sediments. *Geochim. Cosmochim. Acta* **51**, 1897–1921 (1987).

14. Prah, F. G. & Carpenter, R. Hydrocarbons in Washington coastal sediments. *Estuar. Coast. Shelf Sci.* **18**, 703–720 (1984).
15. Vogel, J. S., Nelson, D. E. & Southon, J. R. C-14 background levels in an Accelerator Mass Spectrometry system. *Radiocarbon* **29**, 323–333 (1987).
16. Stuiver, M. & Polach, H. A. Reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).
17. Wakeham, S. G. *et al.* Hydrocarbons in Lake Washington sediments: a 25-year retrospective in an urban lake. *Environ. Sci. Technol.* (in the press).
18. Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J. & Peterson, M. L. Molecular indicators of diagenetic status in marine organic matter. *Geochim. Cosmochim. Acta* **61**, 5363–5369 (1997).
19. Eglinton, T. I. *et al.* Composition, age and provenance of organic matter in NW African dust over the Atlantic Ocean. *Geochem. Geophys. Geosyst.* **3**, 1–27 (2002).
20. Weis, P. L. The origin of epigenetic graphite: evidence from isotopes. *Geochim. Cosmochim. Acta* **45**, 2325–2332 (1981).
21. Schidlowski, M. Carbon isotopes as biogeochemical recorders of life over 3.8 Ga of Earth history: evolution of a concept. *Precamb. Res.* **106**, 117–134 (2001).
22. Petsch, S. T., Smernik, R. J., Eglinton, T. I. & Oades, J. M. A solid-state ^{13}C -NMR study of kerogen degradation during black shale weathering. *Geochim. Cosmochim. Acta* **65**, 1867–1882 (2001).
23. Herring, J. R. in *The Carbon Cycle and Atmospheric CO_2 : Natural Variations, Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 419–442 (AGU, Washington, 1985).
24. Suman, D. O., Kuhlbusch, T. A. J. & Lim, B. in *Sediment Records of Biomass Burning and Global Change* (ed. Clark, J. S.) 271–293 (Springer, Berlin, 1997).
25. Sackett, W. M., Poag, C. W. & Eadie, B. J. Kerogen recycling in the Ross Sea, Antarctica. *Science* **185**, 1045–1047 (1974).
26. Blair, N. E. *et al.* The persistence of memory: the fate of ancient sedimentary organic carbon in a modern sedimentary system. *Geochim. Cosmochim. Acta* **67**, 63–73 (2003).
27. Eglinton, T. I. *et al.* Variability in radiocarbon ages of individual organic compounds from marine sediments. *Science* **277**, 796–799 (1997).
28. Petsch, S. T., Berner, R. A. & Eglinton, T. I. A field study of the chemical weathering of ancient sedimentary organic matter. *Org. Geochem.* **31**, 475–487 (2000).
29. Jahnke, R. A. The global ocean flux of particulate organic carbon: Areal distribution and magnitude. *Glob. Biogeochem. Cycles* **10**, 71–88 (1996).
30. Druffel, E. R. M., Williams, P. M., Livingston, H. D. & Koide, M. Variability of natural and bomb-produced radionuclide distributions in abyssal red clay sediments. *Earth Planet. Sci. Lett.* **71**, 205–214 (1984).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the scientific staff at the Lawrence Livermore National Laboratory Center for Accelerator Mass Spectrometry for assistance with radiocarbon analyses, C. Preston and C. Swanston for radiocarbon analysis of the Stillaguamish River sample, S. Petsch and R. Smernik for providing kerogen samples for analysis, and P. Quay and E. Druffel for editing and comments. This work was supported by a mini-grant from LLNL CAMS and by grants from the NSF; A.E.D. thanks the NSF for a graduate research fellowship; Y.G. thanks the Canadian NSERC and Quebec NATEQ for support for this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.D. (dickensa@u.washington.edu).

Melting of iron at the physical conditions of the Earth's core

Jeffrey H. Nguyen & Neil C. Holmes

Physics and Advanced Technologies, H-Division, Lawrence Livermore National Laboratory, Livermore, California 94551, USA

Seismological data can yield physical properties of the Earth's core, such as its size and seismic anisotropy^{1–3}. A well-constrained iron phase diagram, however, is essential to determine the temperatures at core boundaries and the crystal structure of the solid inner core. To date, the iron phase diagram at high pressure has been investigated experimentally through both laser-heated diamond-anvil cell and shock-compression techniques, as well as through theoretical calculations^{4–17}. Despite these contributions, a consensus on the melt line or the high-pressure, high-temperature phase of iron is lacking. Here we report new and re-analysed sound velocity measurements of shock-compressed iron at Earth-core conditions¹⁵. We show that melting starts at 225 ± 3 GPa ($5,100 \pm 500$ K) and is complete at

260 ± 3 GPa ($6,100 \pm 500$ K), both on the Hugoniot curve—the locus of shock-compressed states. This new melting pressure is lower than previously reported¹⁶, and we find no evidence for a previously reported solid–solid phase transition on the Hugoniot curve near 200 GPa (ref. 16).

The radii of the solid inner core and the molten outer core are roughly 1,220 km (pressure $P = 329$ GPa)¹ and 3,490 km ($P = 136$ GPa), respectively. About 20 years ago, Brown and McQueen reported iron melting at 243 GPa (ref. 16). This has since been the high pressure ‘anchor’ of the iron melt line (Fig. 1). Their proposed solid–solid transition has also driven the search for a solid–solid phase boundary near 200 GPa (refs 4–8). At different times, the γ – ϵ (ref. 4) and β – ϵ (ref. 5) phase boundaries were proposed to explain this solid–solid transition (Fig. 1). However, recent data placed the liquid– γ – ϵ triple point between 50 and 60 GPa (refs 6–8). The β – ϵ phase boundary has been shown to be nearly horizontal, that is, no intersection with the melt curve at 200 GPa (refs 7, 8) (Fig. 1). The existence of the β -phase is also questionable^{6,9}. Thus there is at present no other experimental evidence consistent with the putative solid–solid phase transition reported by Brown and McQueen¹⁶.

To corroborate their sound velocity data¹⁶, we focused our experiments between 200 and 260 GPa, employing similar but improved sound velocity measurement techniques¹⁵. Longitudinal sound velocity (C_L), commonly used to determine melting of metals at high pressures^{15–21}, decreases sharply on melting. In a solid, C_L depends on bulk and shear moduli and density, whereas it is only a function of the bulk modulus and density in a liquid. Upon melting, the shear modulus of the solid–liquid mixture tends to zero causing sound velocity to decrease. In metals such as Ta, Al and Mo, C_L decreases by 5–15% when the Hugoniot crosses the melt curve^{18–21}.

We carried out these experiments using a two-stage gas gun capable of generating pressures as high as 400 GPa in the iron sample. This pressure exceeds that at the centre of the Earth, 361 GPa, making single shock experiments suitable for exploring

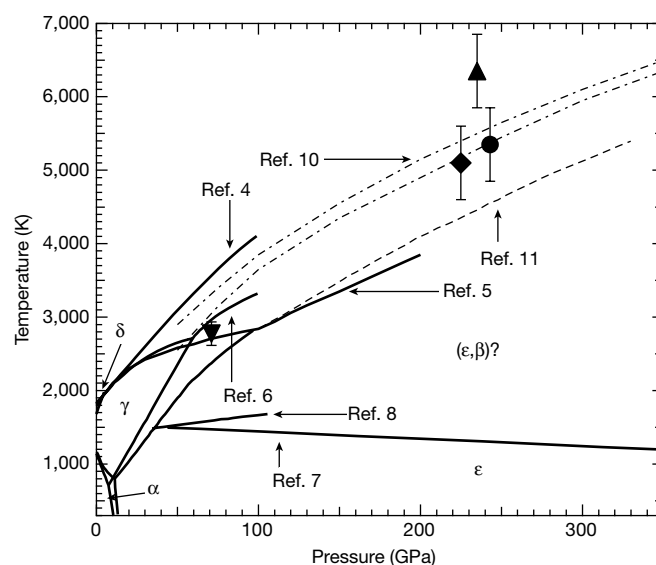


Figure 1 Iron phase diagram. Solid lines, phase boundaries and melt lines from DAC experiments^{4–8}. For clarity, uncertainties of the melt curve, often hundreds of degrees, are not included here. Dashed curve, calculated melt curve by Laio *et al.*¹¹. Upper (lower) dash-dotted curve, calculation with (without) free energy correction by Alfé *et al.*¹⁰. Data points are taken from shock-compression experiments. Triangle data point is from pyrometric temperature measurement by Yoo *et al.*⁹. Circle (Brown and McQueen¹⁹), reverse triangle (Ahrens *et al.*¹⁷) and diamond (this study) data points are from sound velocity experiments. Temperatures of these sound velocity data points are calculated using thermodynamic relations (equation (4)).

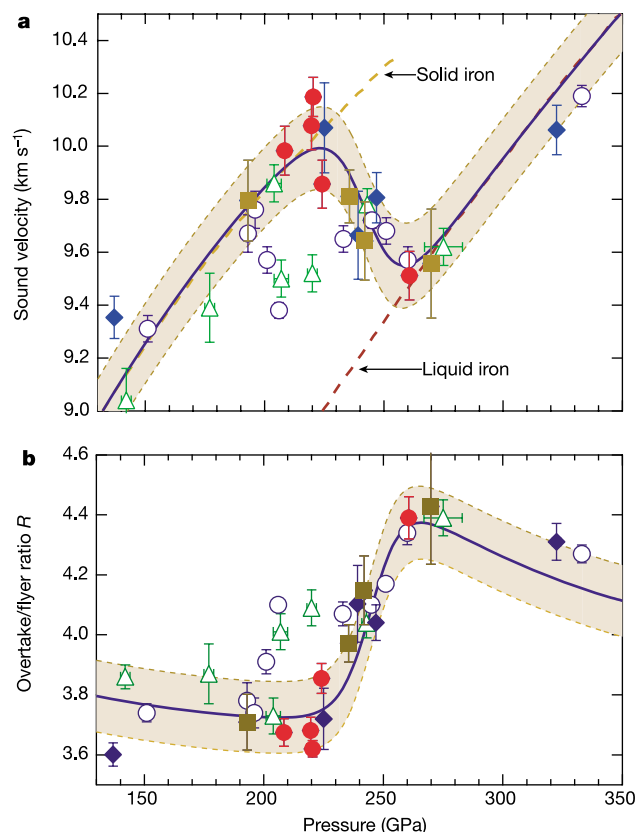


Figure 2 Sound velocity and overtake ratio. **a**, Sound velocity from this study (solid symbols) exhibits a single solid–liquid transition along the Hugoniot at 225 ± 3 GPa. The solid circles (squares, diamonds) represent sample number 1 (2, 3) (see Table 1). Brown and McQueen's¹⁶ data are shown by open symbols. The open circles (triangles) represent gas gun (explosive driven) shock data. The heavy-dashed lines are calculated sound velocities¹⁶. The solid line is a fit to data from this study using equation (2), and the area within the dashed lines is its associated 1σ prediction band. **b**, Overtake ratio of sample plate thickness to impactor plate thickness. This ratio, which is a direct experimental measurement, clearly shows a simple transition. Data are fitted to an equation for R , which is analogous to equation (2). C_L , C_s and C_{liquid} are replaced by R , R_s and R_{liquid} . The latter terms are inverted from the former using equation (1).

core conditions. Just as important, the principal Hugoniot of iron crosses the melt line at a pressure between those of the core–mantle boundary and the inner–outer core boundary. In these single shock experiments, the impact generates two planar shock fronts travelling at shock velocity U_s in opposite directions: one forward into the iron sample, the other backward into the iron impactor. The latter reflects forward, at the back end of the impactor, as a rarefaction wave moving at longitudinal sound velocity C_L . As $C_L + U_p > U_s$ (for a rarefaction wave propagating in shocked material moving at particle velocity U_p)²², the rarefaction wave eventually overtakes the forward-moving shock wave U_s .

As iron is opaque, we measured these overtake events indirectly after the shock and rarefaction fronts have left the iron sample. We designed sample plates with up to seven steps on the non-impact side, and filled the volume behind the stepped surface with bromoform. Above 20 GPa shock pressure, bromoform becomes opaque and emits light whose intensity varies roughly as U_p^8 . This intensity decreases as the rarefaction wave overtakes the shock front (U_p decreases). These two events (shock break-out and rarefaction overtake) define the overtake time. By extrapolating the overtake times to zero, we determined the minimum iron thickness (overtake distance) necessary for an overtake event to take place in iron. Overtake times are a linear function of step thicknesses²³. From the

Table 1 Trace-element concentrations and initial densities

	Sample 1	Sample 2	Sample 3	Ref. 16
C	0.04	0.03		0.02
Al	0.0092	0.0450		
Si	<0.005	0.0060		0.1
P	0.0080	0.0180		
S		0.007		
V	<0.0005	0.0756		
Cr	0.0083	0.0182		
Mn	0.0004	0.3440		0.7
Co	0.0025	0.0039		
Ni	0.0019	0.0148		
Cu	0.0066	0.0040		
ρ_0 (g cm ⁻³)	7.870(3)	7.817(4)	7.859(4)	7.850

Trace element analysis was done by Elemental Research Inc., using laser ablation mass spectroscopy. Carbon analysis was done by LECO, Inc. Nitrogen, oxygen and sulphur trace concentrations were not measured. Sulphur concentration was supplied by the manufacturer. No analysis was done for sample no. 3. Sample densities were measured by the volume displacement method (immersion density). The value in the parentheses represents the uncertainty of the last significant digit.

ratio (R) of overtake distance to impactor thickness (Fig. 2b), we calculated the longitudinal sound velocity as¹⁶:

$$C_L = \left(\frac{\rho_0}{\rho} \right) U_s \frac{(R+1)}{(R-1)} \quad (1)$$

where ρ_0 and ρ are the ambient and shocked iron densities, respectively. The shocked sample density and velocity, ρ and U_s , are calculated from ρ_0 (Table 1) and U_p , using the Hugoniot and conservation relations^{16,22}. In symmetric experiments (iron impactors), particle velocity U_p is exactly one-half that of the impactor velocity²². To make a simple and direct comparison with Brown and McQueen's¹⁶ sound velocity data, we used the same linear fit to the Hugoniot data that they used, $U_s = c + sU_p$ where $c = 3.955 \pm 0.028$ km s⁻¹ and $s = 1.580 \pm 0.011$ (ref. 16). Calculations with their recently revised iron Hugoniot data²⁴ do not affect the measurement of the onset of melting.

Our sound velocity data are in good agreement with that of Brown and McQueen¹⁶, except between 200 GPa and 243 GPa (Fig. 2a). Within this pressure range, we observed a monotonic decrease in C_L and a similar increase in R starting at 225 ± 3 GPa (Fig. 2). In order to discern the number of phase transitions (decreasing steps in sound velocity), we employed two different models: a single solid–liquid phase transition model (equation (2)), and a two-phase transitions model (equation (3)),

$$C_L = [f(P, P_c, \omega)C_\omega] + [(1 - f(P, P_c, \omega))C_{\text{liquid}}] \quad (2)$$

$$C_L = [f(P, P_{c1}, \omega_1)C_\varepsilon] + [(1 - f(P, P_{c1}, \omega_1))f(P, P_{c2}, \omega_2)C_\beta] + [(1 - f(P, P_{c2}, \omega_2))C_{\text{liquid}}] \quad (3)$$

where $f(P, P_c, \omega) = \{1 + \exp[-(P - P_c)/\omega]\}^{-1}$. C_ε , C_β and C_{liquid} are calculated sound velocities for the two proposed solid phases and a liquid phase, respectively¹⁶. $f(P, P_c, \omega)$ is a smoothly varying connecting function of pressure, P , with centre at P_c and width ω ; P_c and ω are used as fitting parameters. C_ε , C_β and C_{liquid} are treated as 'constants' without affecting reduced χ^2 .

The two-phase transition equation (equation (3)) fits well to Brown and McQueen's data¹⁶ (Table 2). Reduced χ^2 results suggest that both models can describe our data (Table 2). However, on closer examination, contribution from the C_β line (second term, equation (3)) seems unconstrained, resulting in an abnormally large 1σ prediction band. On the other hand, the prediction band from equation (2) encompassed approximately 68.3% of our data, as expected (Fig. 2a). From these observations, we conclude that equation (2)—the one phase transition model—sufficiently describes our data. The existence of another solid phase is thus not discernible from our data. This interpretation is consistent with those of Al and Ta^{19–21}.

Table 2 Fitting parameters and reduced χ^2

	Equation (2)	Equation (3)	Ref. 16 (equation (3))
Reduced χ^2	2.68	2.44	4.15
ρ_{c1}	243.5 ± 2.1	224.4 ± 1.1	201.7 ± 0.8
ω_1	7.2 ± 1.8	1.4 ± 2.0	2.6 ± 0.6
ρ_{c2}		252.7 ± 4.3	252.4 ± 1.5
ω_2		7.2 ± 2.4	5.4 ± 1.2

Fitting parameters and reduced χ^2 from equations (2) and (3). Reduced χ^2 is defined as $\chi^2/(N - m)$ where N and m are the numbers of data points and parameters, respectively. For comparison purposes, we fit equation (3) to Brown and McQueen's¹⁶ data in the same pressure range (190 and 275 GPa). Fitting to the entire set of data yields a slightly larger reduced χ^2 , but no significant changes in the fitting parameters.

In Fig. 2a, we segregated our sound velocity data by sample initial densities and concentrations of trace elements (Table 1). Whether taken separately or together, data from all three samples indicate the existence of one single phase transition. Slight differences in sample purity did not affect sound velocity in our study. Whether differences in sound velocity between our samples and those of Brown and McQueen¹⁶ can be attributed to sample purity is not obvious.

Consistent with previous sound velocity studies, we define the onset of melting at the intersection of C_e and the linear extrapolation of the fitted sound velocity at P_c (refs 16, 19–21). We infer from our data a single solid–liquid phase transition starting at 225 ± 3 GPa and completing at 260 ± 3 GPa. The sound velocity decreases by 6% and the pressure range of the solid–melt region extends for 35 ± 4 GPa from 225 ± 3 to 260 ± 3 GPa. This is twice the previous 3% decrease in sound velocity and the 17 ± 3 GPa pressure range for the solid–liquid regime of Brown and McQueen¹⁶. However, this sound velocity decrease falls within range of those of 2024 Al, Mo and Ta (6–14%)^{18–20}. Our Hugoniot pressure range of the mixed region is between those of 2024 Al, Mo and Ta (15–30 GPa)^{18–20} and that of iron (65 GPa)¹³. Thus our interpretation of a single phase transition is consistent with other similar studies^{18–20}.

Temperatures were not measured in these experiments. Instead, they are calculated by integrating the thermodynamic relation,

$$dT = -(T\gamma/V)dV + [(V_o - V)dP + (P - P_o)dV]/(2C_V) \quad (4)$$

along the Hugoniot¹⁶. Accurate temperature calculation requires accurate values of the Grüneisen parameter γ and heat capacity C_V . Only estimates are available; for comparison purposes, we used the same values used by Brown and McQueen¹⁶. Interpolation of their calculations yields a melting temperature of $5,100 \pm 500$ K at 225 GPa, the melting pressure. Wasserman *et al.*'s²⁵ calculations of temperature on the iron Hugoniot, with temperature as a fitting parameter, are also in good agreement with those of Brown and McQueen¹⁶ within cited uncertainty. Interpolation of their calculation yields a temperature of $\sim 5,300$ K at 225 GPa.

Even though our new melting pressure at 225 ± 3 GPa is significantly lower than the previously reported 243 ± 2 GPa (ref. 16), the two calculated temperatures are on approximately the same melt curve (Fig. 1). This was due to the correction for entropy of solid–solid transition at 200 GPa which lowers their melting temperature by roughly 350 K (ref. 16); there was no such correction in our calculation. Our new calculated melting temperature is about $1,000 \pm 700$ K lower than the measured temperature of shock-melted iron¹³, and is higher than the highest measured diamond-anvil cell (DAC) result⁵ by about $1,000 \pm 540$ K. In general, our result falls in between simple extrapolations of DAC melting curves^{4–6}, but assessing uncertainties associated with extrapolation over such a large pressure range is beyond the scope of this work. Calculations by Alfè *et al.*¹⁰, within uncertainties of experiments, match well with sound velocity measurements: this study, Brown and McQueen's¹⁶ and Ahrens *et al.*¹⁷. Laio *et al.*'s¹¹ calculations are more in line with Boehler's DAC results⁵.

Superheating in shock compression experiments has been

observed in non-metals²⁶. However, superheating alone cannot explain the temperature discrepancies between shock and static measurements of melting of transition metals at $P > 100$ GPa (ref. 27). Experimentally, there is no superheating in aluminium²⁸, and evidence is statistically insignificant in iron¹³. This can in part be explained by the fast relaxation times ($\leq 10^{-12}$ s) for electron–electron, electron–phonon and phonon–phonon thermalization in metals²⁹. This fast relaxation time allows equilibrium to be quickly established in shocked metals.

Our observation of no solid–solid phase transition is consistent with both the proposed nearly-horizontal β – ϵ phase boundary^{7,8} and the finding of no β -Fe (ref. 6; Fig. 1). Unfortunately, our experimental pressure range does not cross this new β – ϵ phase boundary and we cannot discern the crystal structure of iron with sound velocity. Nevertheless, existence of a solid–solid phase transition near 200 GPa as previously reported¹⁶ is not consistent with these recent studies^{6–8}—but our finding is consistent.

Using the pure iron melt line to study the temperature of the Earth's core is neither simple nor direct. The Earth's solid inner core is predominantly iron, but available data suggest the possible presence of oxygen and sulphur among others in the liquid outer core. However, with an understanding of eutectic melting depression in the Fe–FeO–FeS system³⁰ at core pressures, we can obtain sufficient estimates for temperature at the core boundaries. □

Received 4 April; accepted 25 November 2003; doi:10.1038/nature02248.

- Lehmann, I. P. *Publ. Bur. Cent. Seismol. Int.* **A 14**, 87–115 (1936).
- Song, X. & Richards, P. G. Seismological evidence for differential rotation of the Earth's inner core. *Nature* **382**, 221–224 (1996).
- Su, W., Dziwowski, A. M. & Jeanloz, R. Planet within a planet: rotation of the inner core of Earth. *Science* **274**, 1883–1887 (1996).
- Williams, Q., Jeanloz, R., Bass, J., Svendsen, B. & Ahrens, T. J. The melting curve of iron to 250 gigapascals: a constraint on the temperature at Earth's center. *Science* **236**, 181–182 (1987).
- Boehler, R. Temperatures in the Earth's core from melting-point measurements of iron at high static pressures. *Nature* **363**, 534–536 (1993).
- Shen, G., Mao, H. K., Hemley, R. J., Duffy, T. S. & Rivers, M. L. Melting and crystal structure of iron at high pressures and temperatures. *Geophys. Res. Lett.* **25**, 373–376 (1998).
- Saxena, S. K. & Dubrovinsky, L. S. Iron phases at high pressures and temperatures: phase transition and melting. *Am. Mineral.* **85**, 372–375 (2000).
- Andraut, D., Fiquet, G., Charpin, T. & Le Bihan, T. Structure analysis and stability field of β -iron at high pressure and temperature. *Am. Mineral.* **85**, 364–371 (2000).
- Yoo, C. S., Akella, J., Campbell, A. J., Mao, H. K. & Hemley, R. J. Phase diagram of iron by in situ x-ray diffraction: implications for Earth's core. *Science* **270**, 1473–1475 (1995).
- Alfè, D., Price, G. D. & Gillan, M. J. Iron under Earth's core conditions: Liquid-state thermodynamics and high-pressure melting curve from ab initio calculations. *Phys. Rev. B* **65**, 165118 (2002).
- Laio, A., Bernard, S., Chiarotti, G. L., Scandolo, S. & Tosatti, E. Physics of iron at Earth's core conditions. *Science* **287**, 1027–1030 (2000).
- Ahrens, T. J., Bass, J. D. & Abelson, J. R. in *Shock Compression of Condensed Matter—1989* (eds Schmidt, S. C., Johnson, L. W. & Davison, L. W.) (Elsevier Science, Amsterdam, 1990).
- Yoo, C. S., Holmes, N. C., Ross, M., Webb, D. J. & Pike, C. Shock temperatures and melting of iron at Earth core conditions. *Phys. Rev. Lett.* **70**, 3931–3934 (1993).
- Benuzzi-Mounaix, A. *et al.* Absolute equation of state measurements of iron using laser driven shocks. *Phys. Plasma* **9**, 2466–2469 (2002).
- Nguyen, J. H. & Holmes, N. C. in *Shock Compression of Condensed Matter—1999* (eds Furnish, M. D., Chhabildas, L. C. & Hixson, R. S.) 81–84 (American Institute of Physics, New York, 1999).
- Brown, J. M. & McQueen, R. G. Phase transitions, Grüneisen parameter, and elasticity for shocked iron between 77 GPa and 400 GPa. *J. Geophys. Res.* **91**, 7485–7494 (1986).
- Ahrens, T. J., Holland, K. G. & Chen, C. Q. Phase diagram of iron, revised-core temperatures. *Geophys. Res. Lett.* **29**, 54–54–4 (2002).
- Hixson, R. S., Boness, D. A., Shane, J. W. & Moriarty, J. A. Acoustic velocities and phase transitions in molybdenum under strong shock compression. *Phys. Rev. Lett.* **62**, 637–640 (1989).
- Brown, J. M. & Shaner, J. W. in *Shock Compression of Condensed Matter—1983* (eds Asay, J. R., Graham, R. A. & Straub, G. K.) 91–94 (North-Holland, Amsterdam, 1984).
- McQueen, R. G., Fritz, J. N. & Morris, C. E. in *Shock Compression of Condensed Matter—1983* (eds Asay, J. R., Graham, R. A. & Straub, G. K.) 95–98 (North-Holland, Amsterdam, 1984).
- Shaner, J. W., Brown, J. M. & McQueen, R. G. Melting of metals above 100 GPa. *Mater. Res. Soc. Symp. Proc.* **22**, 137–141 (1984).
- Zel'dovich, Y. B. & Raizer, Y. P. *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Academic, New York, 1967).
- McQueen, R. G., Hopson, J. W. & Fritz, J. N. Optical technique for determining rarefaction wave velocities at very high pressures. *Rev. Sci. Instrum.* **53**, 245–250 (1982).
- Brown, J. M., Fritz, J. N. & Hixson, R. S. Hugoniot data for iron. *J. Appl. Phys.* **88**, 5496–5498 (2000).
- Wasserman, E., Stixrude, L. & Cohen, R. E. Thermal properties of iron at high pressures and temperatures. *Phys. Rev. B* **53**, 8296–8309 (1996).
- Lou, S. N. & Ahrens, T. J. Superheating systematics of crystalline solids. *Appl. Phys. Lett.* **82**, 1836–1838 (2003).
- Lou, S. N. & Ahrens, T. J. Application of shock-induced superheating to the melting of geophysically important minerals. *Phys. Earth Planet. Inter.* (in the press).

28. Boehler, R. & Ross, M. Melting curve of aluminum in a diamond cell to 0.8 Mbar: implications for iron. *Earth Planet. Sci. Lett.* **153**, 223–227 (1997).
29. Wallace, D. C. Irreversible thermodynamics of flow in solids. *Phys. Rev. B* **22**, 1477–1486 (1980).
30. Boehler, R. Melting of the Fe–FeO and Fe–FeS systems at high-pressure—constraints on core temperatures. *Earth Planet. Sci. Lett.* **111**, 217–227 (1992).

Acknowledgements We benefited from discussions with J. M. Brown, O. L. Anderson, M. Ross and R. Boehler. We acknowledge F. H. Streitz for the formulation of equations (2) and (3). We are grateful for the technical efforts of S. Caldwell, E. Ojala, L. Raper, K. Stickle. Work was performed by the University of California under the auspices of the US DOE by the Lawrence Livermore National Laboratory.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H.N. (nguyen29@llnl.gov).

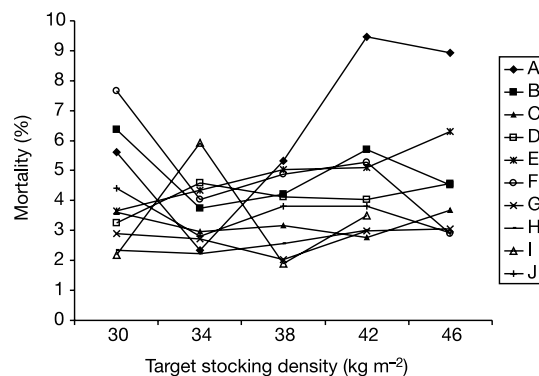


Figure 1 Total mortality in relation to target stocking density. The percentage of total mortality is shown separately for each company (A–J) taking part in the trial.

Chicken welfare is influenced more by housing conditions than by stocking density

Marian Stamp Dawkins, Christi A. Donnelly* & Tracey A. Jones

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

* Present address: Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, St. Mary's Campus, Norfolk Place, London W2 1PB, UK

Intensive broiler (meat) chicken production now exceeds 800 million birds each year in the United Kingdom and 2×10^{10} birds worldwide¹, but it attracts accusations of poor welfare^{2,3}. The European Union is currently adopting standards for broilers aimed at a chief welfare concern—namely, overcrowding—by limiting maximum ‘stocking density’ (bird weight per unit area). It is not clear, however, whether this will genuinely improve bird welfare because evidence is contradictory^{4–10}. Here we report on broiler welfare in relation to the European Union proposals through a large-scale study (2.7 million birds) with the unprecedented cooperation of ten major broiler producers in an experimental manipulation of stocking density under a range of commercial conditions. Producer companies stocked birds to five different final densities, but otherwise followed company practice, which we recorded in addition to temperature, humidity, litter and air quality. We assessed welfare through mortality, physiology, behaviour and health, with an emphasis on leg health and walking ability. Our results show that differences among producers in the environment that they provide for chickens have more impact on welfare than has stocking density itself.

Across companies, there was a wide range of house sizes (455–1,901 m²), house ages (5–40 yr) and numbers of birds per house (7,500–53,000); 75% of flocks were Ross 308 (see Supplementary Information). Each company contributed two houses to each target stocking density for a trial. With two companies, we repeated trials with the same ten houses in summer and winter to examine the effects of season. Within a company, houses were randomly assigned to stocking density (Supplementary Information), which was manipulated by altering the numbers of day-old chicks placed to achieve a projected ‘target’ maximum stocking density just before the birds were killed (39–42 d at 2–3 kg). The five target stocking densities were 30, 34, 38, 42 and 46 kg m^{−2} (refs 1, 11, 12). Actual stocking density was measured as mean weight $\times$ number of birds per area of house. The same person (T.A.J.) made or checked all measurements with the help of trained assistants.

Welfare^{13–15} was assessed through mortality, physiology, behaviour and health, emphasizing leg health and walking ability^{16–20} (Table 1). We found that the effect of experimentally manipulating stocking density was overshadowed by much larger differences among companies (Table 2 and Fig. 1). Chickens grew more slowly at the highest stocking densities and jostled each other more, and fewer of them showed the best gaits (Table 3); however, for the most obvious measures of bird welfare—that is, the numbers of birds dying, being culled as unfit and showing leg defects—there was no effect of stocking density. There were, however, substantial differences among companies in almost all measures examined (Table 2). At no point was breed a significant explanatory factor in any outcome variables, suggesting that the differences were due to environmental influences.

Of the commercially relevant factors that seemed to allow some companies to ‘cope’ better than others with high stocking densities, the most likely candidates were those that affected litter moisture

Table 1 Scoring of gait, hockburn, pad dermatitis and leg deviations

Leg health measure	Score 0	Score 1	Score 2
Gait	Bird walks with ease, has regular and even strides and is well balanced	Bird walks with irregular and uneven strides and appears unbalanced	Bird is reluctant to move and is unable to walk many strides before sitting down
Hockburn*	No discolouration or lesions	<10% hock with lesion	>10% hock with lesion
Pad dermatitis†	No lesions	<5 mm lesion on pad	>5 mm lesion on pad
Angle: in	Legs straight	Inward bow at intertarsal joint so that the two legs meet >22°	
Angle: out	Legs straight	Outward twist at intertarsal joint with ≥30° between the legs	
Rotation	Legs straight, pads facing away from handler	Rotation of the tibia shaft so that pads face each other >15°	

* Pink hocks were also recorded.

† Pervasively dirty pads also scored.

Table 2 Density and company effects on principal parameters

	Mean	s.e.m.	Range	Target density	Company	Actual density
Gait score 0 (%)	72.6	2.0	10–100	$P = 0.027$	$P < 0.0001$	$P = 0.002$
Gait score 2 (%)	0.9	0.3	0–20	NS*	NS	NS
Hock score 0 (%)	80.6	1.7	20–100	NS	$P < 0.0001$	NS
Hock score 2 (%)	1.5	0.5	0–47.5	NS	$P = 0.0003$	NS
Pad score 0 (%)	81.2	2.1	12.5–100	NS	$P < 0.0001$	NS
Pad score 2 (%)	2.8	0.7	0–47.5	NS	NS	NS
Angle: in (%)	12.3	0.7	0–35	NS	$P < 0.0028$	NS
Angle: out (%)	3.5	0.4	0–17.5	NS	$P < 0.0001$	NS
Rotation (%)	6.9	0.7	0–47.5	NS	$P < 0.0001$	NS
Total leg deviation (%)	23.3	1.2	0–62.5	NS	$P < 0.0001$	NS
Corticosteroid (ng ml ⁻¹)	18.9	1.1	3.5–50.4	NS	$P < 0.0001$	NS
Growth rate (g d ⁻¹)	49.4	0.4	39.1–7.9	$P = 0.011$	$P < 0.0001$	NS
Total mortality (%)	4.1	0.2	1.4–14.7	NS	$P < 0.0001$	NS
Leg cull (%)	0.6	0.1	0–2.4	NS	$P < 0.0001$	NS
Other cull (%)	1.5	0.1	0.4–4.7	NS	$P < 0.0001$	NS
Dead birds (%)	2.0	0.1	0.6–4.8	NS	$P < 0.0001$	NS
Downgrades (%)	0.9	0.1	0–4.24	NS	$P < 0.0001$	NS
Litter moisture (%)	18.3	0.4	10.7–2.5	$P = 0.05$	$P < 0.0001$	NS
Ammonia (p.p.m.)	10.4	0.7	1.3–29.8	NS	$P < 0.0001$	NS

*NS, not significant.

Table 3 Gait 0, jostle and growth rates* for given target stocking densities

	30 kg m ⁻²	34 kg m ⁻²	38 kg m ⁻²	42 kg m ⁻²	46 kg m ⁻²
Gait score 0 (%)	80.8 ^a (3.7)	74.2 ^{abc} (4.9)	76.1 ^{ac} (4.3)	68.0 ^{bc} (4.2)	61.1 ^{bc} (4.5)
Jostle rate (incidents per min)	0.316 ^a (0.039)	0.431 ^{ab} (0.046)	0.455 ^{ab} (0.049)	0.566 ^b (0.071)	0.618 ^b (0.112)
Growth rate (g d ⁻¹)	50.3 ^a (0.8)	49.9 ^a (0.8)	49.7 ^{ab} (0.9)	48.8 ^{ab} (0.9)	47.7 ^b (0.9)

*Values are the mean, the s.e.m. is given in parentheses. Values with different superscripts are significantly different, $P \leq 0.05$ (Tukey post hoc comparison).

and air ammonia, because these differed significantly among companies (Table 2). Whereas 56% of the variation in litter moisture could be explained by effects such as heater position in the house and the number of drinkers per 1,000 birds, 73.3% of the variation in air ammonia could be explained by effects such as season and ventilation type.

Litter moisture and ammonia, in turn, were related to bird health. Higher levels of both were correlated with more dirty pads (moisture, correlation coefficient, $r = 0.24$; ammonia, $r = 0.27$), more legs scored as angle-out (moisture, $r = 0.24$; ammonia, $r = 0.36$) and fewer birds with unblemished hocks (moisture, $r = -0.36$, ammonia, $r = -0.26$). In addition, birds had more hock lesions with wetter litter (score 2, $r = 0.27$). High concentrations of ammonia²¹ were, unexpectedly, associated with lower mortality, but both litter moisture and air ammonia were correlated with higher faecal corticosteroid (a 'stress' hormone^{22,23}; see Supplementary Information). Eighty-four per cent of the variation in faecal corticosteroid could be explained by effects such as temperature in week 1, humidity in week 5, season and ventilation type. Corticosteroid concentrations were also correlated with mortality (Supplementary Information), which suggests that stress on birds and their risk of dying depend on the extent to which companies can control the house environment.

The key house variables were temperature and humidity. The percentage of birds dying over the whole growth period was positively correlated with humidity and temperature in weeks 3–5 (Supplementary Information). The number of birds walking well (gait 0) was negatively correlated with the percentage of time that temperature and humidity were outside the breeder's recommended range²⁴ in weeks 1 and 4. All companies deliberately altered house temperature²⁴ as the birds grew: 84.6% of that variation was explained by week, company (which included litter type and number of stockman visits), and the covariates drinker type, season and numbers of drinkers per 1,000 birds. None of the companies measured or controlled humidity directly, but this also changed as the birds grew. Week, company (including the number of stockman visits, position of heaters, ventilation type and age of house) and the

number of drinkers per 1,000 birds explained 71.2% of variation in humidity.

Differences between summer and winter (when ventilation is reduced to conserve heat) confirm the importance of house environment. Both mortality and corticosteroid concentrations were lower in summer than in winter (mean summer mortality, 4.2%; mean winter, 5.3%; mean summer corticosteroid, 8.6 ng ml⁻¹; mean winter, 36.2 ng ml⁻¹). Foot pad condition was correspondingly better with more birds scoring pad 0 in summer (88.9%) than winter (71.6%) and fewer birds having pervasively dirty pads (8.8% in summer versus 30.3% in winter).

Although house environment is crucial to bird welfare, we emphasize that stocking density is also important. At the two highest target stocking densities (42 and 46 kg m⁻²), there were fewer birds with the best gait score 0. Culling rates were not greater (Table 2), which suggests that at least some aspects of leg health are compromised at or above a stocking density of 42 kg m⁻².

We conclude that, although very high stocking densities do affect chicken welfare, stocking density *per se* is, within limits, less important than other factors in the birds' environment that differ among companies. Good stockmanship counts even in highly automated houses. Legislation to limit stocking density that does not consider the environment that the birds experience could thus have major repercussions for European poultry producers without the hoped-for improvements in animal welfare. These will come from improving the environment²⁵, nutrition²⁶ and genetics^{27,28} of the millions of birds that we eat. □

Methods

Management and husbandry

We recorded the following information: house age; ownership (company or contract grower); size; orientation; fabric; light pattern and source; dawn/dusk dimming; percentage of wheat feed; feeder type and number; drinker type and number; heater type, number and position; type of ventilation system; misting systems; floor and litter type; number of stockmen and number of visits per house; vaccination programme; feed withdrawal; and thinning or clearance programme.

Environment

We recorded temperature and relative humidity every hour throughout the growth cycle

(by using four randomly positioned Tiny Talk data loggers in each house at a height of 60 cm); and atmospheric ammonia (by a Gastec GV-100S pumpset), light at bird height (ISO-Tech digital light meter) and litter moisture content ([sample weight difference after drying/sample weight] $\times$ 100) at target density.

Birds

We recorded the source, breed, sex (all-male, all-female or mixed), age of parent flock, date and time of arrival, position of chicks on delivery lorry, type of vehicle (rigid or articulated), ventilation (fans, vents) and on-board temperature. The numbers of trays of chicks per house and chicks per tray were audited.

Leg health

At target density, a single bird chosen at random from each of ten points in each house, was observed walking for at least ten paces before being scored for gait (Table 1, $n = 1,140$ birds). Groups of ten birds were subsequently caught at four random points per house; individuals were inverted (ventral side facing handler), held by the legs with the handler's thumbs just below the intertarsal joint and assessed for leg straightness (Table 1, $n = 4,370$), weighed and then released.

Corticosteroid measurements

Fresh faecal samples were collected at five random positions in each house, dried at 40 °C and analysed for corticosterone^{19,20}.

Behaviour

Four battery-operated video cameras radio-linked to a VCR (Tracksys) were placed in each house at a height of 155 cm. Eight 10-min sequential records of each camera view were made between 10:00 and 12:00 at target density. One randomly chosen focal bird from each 10-min section of video was analysed for 5 min for frequency and duration of stand, lie, feed, drink, preen, rest (eyes closed) and lie stretched out; frequency of walk (including number of strides) and peck litter, peck other bird, scratch litter, scratch head, stretch head, wing or leg, shake body, shake head, dust bathe, wing flap, aggressive interactions and perch; changes of posture (up or down); jostling or being jostled by other birds; and being disturbed or walked on by other birds ($n = 741$ from 107 houses).

Production

We recorded mortality (numbers of birds found dead plus numbers of birds culled because of illness or leg problems), feed conversion ratio, water intake, date, numbers and weights of birds removed from the house (thinned or cleared) and number of birds rejected at the processing plant. Growth rate was calculated as: individual weight(average chick weight)/number of days.

Statistical analysis

The independent statistical unit was house. Where many measurements were made per house, a single mean-per-house value was used in the analysis. Variables were first analysed for effects of target stocking density, actual stocking density and company by analysis of variance. Where actual density effects were significant, they were further analysed by regression analysis (fitted line model) and post-hoc Tukey comparison.

Univariate linear correlations were examined between outcome variables and predictors treated as continuous variables. Multivariate linear models were constructed using a stepwise model selection procedure (starting from a model with no predictors) with possible predictors, including those continuous predictors with substantial linear correlations (< -0.2 or > 0.2) and categorical predictors. We discuss only effects where $P < 0.01$.

Received 26 June; accepted 19 November 2003; doi:10.1038/nature02226.

1. Scientific Committee on Animal Health and Animal Welfare *The Welfare of Chickens Kept for Meat Production (Broilers)* (European Commission, 2000).
2. Webster, J. *Animal Welfare: a Cool Eye Towards Eden* (Blackwell Scientific, Oxford, 1994).
3. Stevenson, P. *The Welfare of Broiler Chickens* (Compassion in World Farming, Petersfield, UK, 1995).
4. Bagshaw, C. S. & Matthews, L. R. *Broiler Welfare—a Review of Latest Research and Projects in Progress Internationally* (Ministry of Agriculture and Forestry, New Zealand, 2001).
5. Hall, A. The effect of stocking density on the welfare and behaviour of broiler chickens reared commercially. *Anim. Welf.* **10**, 23–40 (2001).
6. Grashorn, M. & Kutritz, B. Effect of stocking density on performance of modern broiler breeds. *Arch. Geflügelkd.* **55**, 84–90 (1991).
7. McLean, J. A., Savory, C. J. & Sparks, N. H. C. Welfare of male and female broiler chickens in relation to stocking density, as indicated by performance, health and behaviour. *Anim. Welf.* **11**, 55–73 (2002).
8. Sørensen, P., Su, G. & Kestin, S. C. Effects of age and stocking density on leg weakness in broiler chickens. *Poult. Sci.* **79**, 864–870 (2000).
9. Feddes, J. J. R., Emmanuel, E. J. & Zidhoff, M. J. Broiler performance, body weight variance, feed and water intake and carcass quality at different stocking densities. *Poult. Sci.* **81**, 774–779 (2002).
10. Heier, B. T., Hogasen, H. R. & Jarp, J. Factors associated with cumulative mortality in Norwegian broiler flocks. *Prev. Vet. Med.* **53**, 147–165 (2002).
11. Farm Animal Welfare Council *Report on the Welfare of Broiler Chickens* (FAWC, Tolworth, UK, 1992).
12. Royal Society for the Prevention of Cruelty to Animals *The RSPCA Welfare Standards for Chickens* (RSPCA, Horsham, UK, 1995).
13. Broom, D. M. Indicators of poor welfare. *Br. Vet. J.* **142**, 524–526 (1986).
14. Dawkins, M. S. in *Coping with Challenge: Welfare in Animals including Humans* (ed. Broom, D. M.) 63–78 (Dahlem Univ. Press, Berlin, 2001).
15. Mendl, M. Assessing the welfare state. *Nature* **410**, 31–32 (2001).
16. Martrenchar, A., Boilletot, E., Huoinic, D. & Pol, F. Risk factors for foot-pad dermatitis in chicken and turkey broilers in France. *Prev. Vet. Med.* **52**, 213–226 (2002).

17. Ekstrand, C., Carpenter, T. E. & Andersson, I. Prevalence and prevention of foot pad dermatitis in broilers in Sweden. *Br. Poult. Sci.* **39**, 318–324 (1998).
18. Kestin, S. C., Gordon, S., Su, G. & Sørensen, P. Relationships in broiler chickens between lameness, live-weight growth rate and age. *Vet. Rec.* **148**, 195–197 (2001).
19. Bradshaw, R. H., Kirkden, R. D. & Broom, D. M. A review of the aetiology and pathology of leg weakness in broilers in relation to welfare. *Avian Poult. Biol. Rev.* **13**, 45–103 (2002).
20. Weeks, C. A., Danbury, T. D., Davies, H. C., Hunt, P. & Kestin, S. C. The behaviour of broiler chickens and its modification by lameness. *Appl. Anim. Behav. Sci.* **67**, 111–125 (2000).
21. Kristensen, E. K. M. & Wathes, C. M. Ammonia and poultry: a review. *World's Poult. Sci. J.* **56**, 235–245 (2000).
22. Cockrem, J. F. & Rounce, J. R. Faecal measurements of oestradiol and testosterone allow non-invasive estimation of plasma steroid concentrations in the domestic fowl. *Br. Poult. Sci.* **35**, 433–443 (1994).
23. Denhard, M. *et al.* Measurement of plasma corticosterone and fecal glucocorticoid metabolites in the chicken (*Gallus domesticus*), the great cormorant (*Phalacrocorax carbo*), and the goshawk (*Accipiter gentilis*). *Gen. Comp. Endocrinol.* **131**, 345–352 (2003).
24. Kemp, C. & Kenny, M. *Ross Broiler Management Manual* 36 (Aviagen, Newbridge, UK, 2002).
25. Reiter, K. & Bessei, W. Möglichkeiten zur Verringerung von Beinschaden bei Broilern und Puten (Übersicht). *Arch. Geflügelkd.* **62**, 145–149 (1998).
26. Whitehead, C. C. Nutrition and poultry welfare. *World's Poult. Sci. J.* **58**, 349–356 (2002).
27. Kestin, S. C., Knowles, T. G., Tinch, A. F. & Gregory, N. G. The prevalence of leg weakness in broiler chickens and its relationship with genotype. *Vet. Rec.* **131**, 190–194 (1992).
28. Gordon, S. H., Charles, D. R. & Green, G. Metabolic age: a basis for comparison of traditional breeds of meat chickens. *Br. Poult. Sci.* **42**, S118–S119 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank Banham Poultry Ltd, Dove Valley (Ashbourne) Ltd, Faccenda Group (including Webb Country Food Group), Grampian Country Chicken, L&M Food Group, Moy Park Ltd, O'Kane Poultry Ltd, G. W. Padley Poultry Ltd, Rose Brand Poultry (including Skovsgaard) and Two Sisters Ltd (including Premier Farming) for participation; and P. Harvey for comments on the manuscript. We thank Defra for funding. M.S.D. conceived the project, made links with the companies, obtained funding, designed the protocols, ran the first trial, took part in about 25% of subsequent trials and largely wrote the paper. C.A.D. advised on the experimental design and undertook much of the statistical analysis. T.A.J. took over the running of the project from trial 2, collected and collated data from all subsequent trials, did the preliminary statistical analysis and collaborated in writing the paper.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to M.S.D. (marian.dawkins@zoo.ox.ac.uk).

Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand

Derek A.T. Cummings^{1,2}, Rafael A. Irizarry³, Norden E. Huang⁴, Timothy P. Endy⁵, Ananda Nisalak⁶, Kumnuan Ungchusak⁷ & Donald S. Burke²

¹Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA

²Department of International Health, and ³Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA

⁴Laboratory for Hydrospheric Processes/Oceans and Ice Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

⁵Virology Division, United States Army Medical Research Institute in Infectious Disease, Fort Detrick, Maryland 21702, USA

⁶Department of Virology, Armed Forces Research Institute of Medical Sciences, Bangkok 10400, Thailand

⁷Bureau of Epidemiology, Ministry of Public Health, Nonthaburi 11000, Thailand

Dengue fever is a mosquito-borne virus that infects 50–100 million people each year¹. Of these infections, 200,000–500,000 occur as the severe, life-threatening form of the disease, dengue haemorrhagic fever (DHF)². Large, unanticipated epidemics of DHF often overwhelm health systems³. An understanding of the spatial-temporal pattern of DHF incidence would aid the

allocation of resources to combat these epidemics. Here we examine the spatial-temporal dynamics of DHF incidence in a data set describing 850,000 infections occurring in 72 provinces of Thailand during the period 1983 to 1997. We use the method of empirical mode decomposition⁴ to show the existence of a spatial-temporal travelling wave in the incidence of DHF. We observe this wave in a three-year periodic component of variance, which is thought to reflect host-pathogen population dynamics^{5,6}. The wave emanates from Bangkok, the largest city in Thailand, moving radially at a speed of 148 km per month. This finding provides an important starting point for detecting and characterizing the key processes that contribute to the spatial-temporal dynamics of DHF in Thailand.

The incidence of DHF in Thailand varies widely from year to year, showing as much as a tenfold difference between years. Dengue is a leading cause of hospitalization and death among children in Thailand, where all four serotypes of the virus circulate⁷. Few tools exist to control dengue virus infection and transmission. Control efforts focus on controlling the mosquito vector of the disease, *Aedes aegypti*, and on effective management of cases of infections¹. Reliable prediction of the location and times of high incidence would allow public health systems to allocate their limited resources more effectively.

It is difficult to predict the pattern of DHF over time and geography (that is, the spatial-temporal pattern) owing to the presence of nonstationarity and nonlinearity in incidence data. Several factors are thought to influence the pattern of DHF, including environmental and climate factors^{8,9}, predator-prey dynamics between the pathogen and the host population^{5,6} and viral factors^{10,11}. Incidence patterns reflect the complex interaction

of all of these factors. As a result, incidence data show strong seasonality, multiyear oscillations and changes in period over time. In the present analysis, we use empirical mode decomposition (EMD) to isolate a 3-yr periodic mode of variance. The travelling wave revealed in this periodic mode is obscured in the raw incidence data by the presence of many periodic and roughly periodic components.

An array of spatial-temporal patterns have been observed^{12–14} and are predicted by theory^{15,16} for host-pathogen and predator-prey ecological systems. A repeating, spatial-temporal wave has not been observed, however, in a vector-borne disease of humans. Characterization of the particular spatial-temporal pattern of an ecological system can be used to identify the mechanisms most important in the dynamics of these systems¹⁷. One feature of disease systems that has been shown to produce waves in incidence is spatial heterogeneity in the host population¹². Spatial temporal travelling waves in measles incidence have been attributed to the reintroduction of measles to small communities through infective sparks from larger communities¹². The size of Bangkok's population and its large role in the commerce of the country suggest that, if spatial heterogeneity in the host population is important to DHF dynamics, Bangkok may have a central role.

The method of EMD is analogous to other methods available for processing nonstationary data, such as wavelet analysis and singular spectrum analysis (SST). Unlike wavelet analysis, however, it does not assume a basis *a priori*. Unlike both wavelet analysis and SST, EMD is appropriate for data describing nonlinear phenomena⁴. EMD uses an adaptive basis that is derived from each data set to decompose the variance of that set into a finite number of intrinsic mode functions (IMFs)⁴. These IMFs represent oscillations around a mean at a characteristic timescale of the data—the spacing between extrema. The IMFs are not restricted to a particular frequency or band of frequency, but can experience both amplitude and frequency modulation. An example of the EMD sifting process is shown in Fig. 1. Features suggesting nonlinearity, such as front-back asymmetry in the waveforms, can be seen in the IMFs presented. The decomposition is local, complete and, for practical purposes (but not theoretically), orthogonal⁴.

Figure 2 shows the monthly incidence of DHF for each Thai province. Brief inspection suggests that there is temporal synchrony across the country, with peaks in incidence occurring roughly at the same time across all provinces. Figure 3 shows the IMFs of roughly

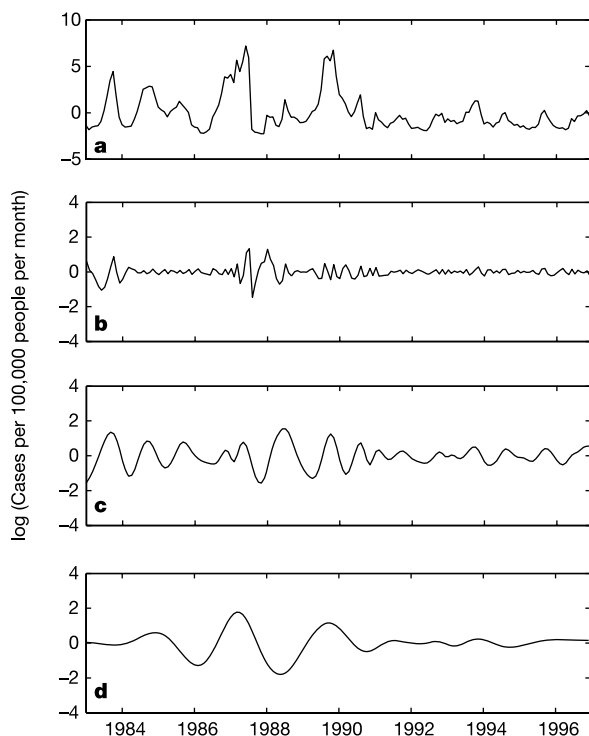


Figure 1 Example of the EMD sifting process. **a**, Time series of the monthly DHF incidence in Bangkok, 1983–1997; see Methods. **b**, Time series of the highest frequency IMF. **c**, Time series of the seasonal IMF. **d**, Time series of the 3-yr periodic IMF. The last three IMFs are not shown; these modes are very slow changing modes representing trends of over 10 yr.

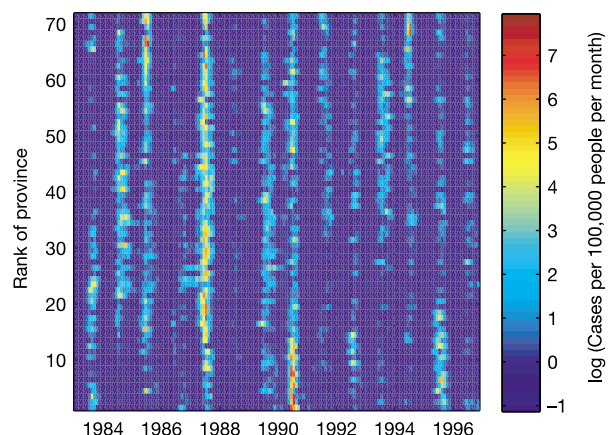


Figure 2 Monthly DHF incidence in each of the 72 provinces of Thailand. Data have been log-transformed and normalized to zero mean and unit variance. Data are presented for provinces from the most southerly to the most northerly from bottom to top. There are 12,096 individual data points.

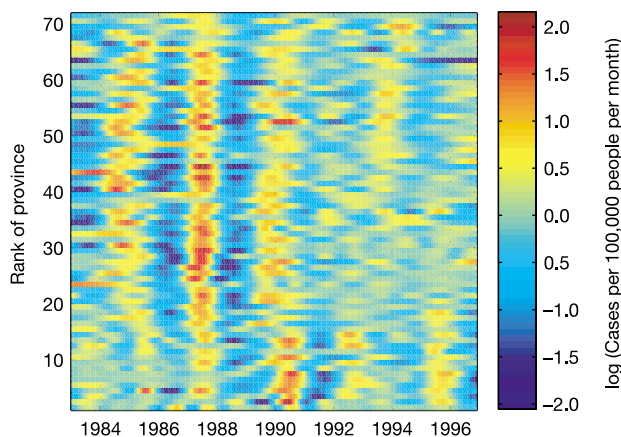


Figure 3 The 3-yr periodic mode for each of the 72 provinces of Thailand. Data are presented for provinces from the most southerly to the most northerly from bottom to top.

3-yr periodicity for all provinces. EMD analysis of each incidence time series yields, at most, six IMFs. The two most energetic IMFs across all provinces are the IMF associated with seasonal variance and the 3-yr periodic IMF. The 3-yr periodic modes account for 44% (95% confidence interval (C.I.) 39–48%) of the interannual variability in dengue incidence.

The nonparametric covariance function¹⁸ was used to characterize the spatial synchrony of incidence fluctuations. Spatial synchrony provides a measure of the spatial dependence of temporal correlation among incidence series. Spatial synchrony of the 3-yr mode differs markedly from the synchrony of the raw data. The spatial extent, the distance for which local synchrony is statistically significantly different from the average synchrony across all data, and the global synchrony of these two sets differ greatly (Fig. 4). The spatial extent of synchrony in the raw DHF incidence is about 180 km, whereas the extent of synchrony of the 3-yr mode is about 420 km.

Spatial synchrony reflects both the timing and relative amplitude of incidence across provinces. By contrast, phase coherence measures only the relative timing of peaks and troughs in incidence¹². Phase coherence of the 3-yr mode (see Supplementary Information), although showing a similar spatial extent, is significantly less than spatial synchrony. This suggests that the timing of peaks and troughs is synchronous during large amplitude changes in incidence, but less synchronous during periods of low incidence. The first half of the data, a period of high energy in the 3-yr mode, has both higher spatial synchrony and coherence than the second half.

To examine the role of Bangkok in this wave, cross-correlation functions (CCFs) between the 3-yr mode of incidence in Bangkok and all other provinces were calculated. Figure 5 shows a marked pattern in these functions. The lag at which each of these CCFs are at their maximum absolute value is found to be a function of the distance from Bangkok ($P < 10^{-8}$; see Supplementary Information). Out of 71 provinces, 68 are either synchronous or lag behind Bangkok. These results describe a repeating, spatial-temporal travelling wave, emanating from Bangkok at a speed of 148 km per month (95% C.I., 114–209 km month⁻¹).

This travelling wave is surprising in its spatial extent. The spatial extent of synchrony of this mode (420 km) and the large number of provinces that are either synchronous or lag behind Bangkok suggest that almost all of the country is affected by this wave. But several border regions do not correlate well with Bangkok. Future research will investigate these border provinces

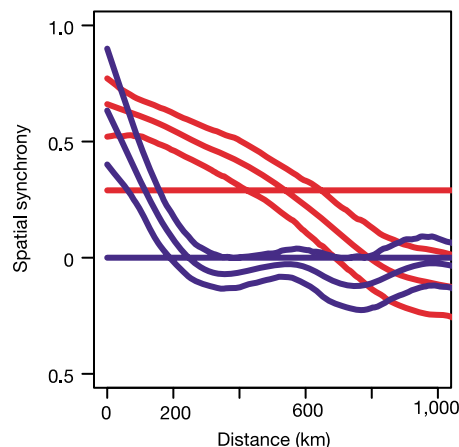


Figure 4 Spatial synchrony of DHF incidence (blue) and the 3-yr periodic mode of variance (red) across 72 provinces of Thailand with 95% C.I. envelopes (see Methods). Horizontal lines indicate the global synchrony for each series.

and whether other urban areas in Southeast Asia influence the incidence patterns in these provinces.

To test the robustness of the central role of Bangkok in the dynamics of DHF in Thailand, we repeated the analysis presented here for all 72 provinces. Bangkok has the largest number of provinces lagging behind its incidence pattern (65/71). Three other provinces tie with Bangkok in having the same number of provincial incidence patterns that are either synchronous or lagging behind their own incidence patterns. In a direct comparison with Bangkok, all three of these provinces lag behind Bangkok's incidence pattern.

The mechanisms underlying this wave are not understood at present. Several classes of mechanism have been implicated in inducing spatial synchrony in other systems, including dispersal^{12,19}, activator–inhibitor dynamics²⁰ and wide-scale correlation of environmental factors²¹. For the system studied here, DHF in Thailand, we speculate that immune interactions between the four serotypes may give rise to complex local dynamics in Bangkok⁵; it has been reported that all four serotypes are continuously present but vary in proportion from season to season²². Similar to the observations for measles in the United Kingdom¹², we speculate that smaller communities elsewhere in Thailand may experience periods of no incidence of particular serotypes during some periods ('stochastic fade-out'), and that these serotypes are reintroduced by the migration of viruses from Bangkok. A preliminary analysis has found that the number of months of zero incidence that each province experiences is associated with population size and distance from Bangkok. A previous study detected no role for weather or climate in longer term variances in DHF incidence in Bangkok⁶; thus, it is unlikely that changes in weather contribute to the country-wide travelling wave.

The development of spatial transmission models of DHF to explore the role of these mechanisms for dengue is an area for future research. Although only narrowing the field of theoretical models¹⁷, the spatial-temporal pattern described here provides a much needed crucial test for models of the spatial transmission of dengue. The 3-yr periodic mode seems to modulate both in frequency and in amplitude. It is possible that this modulation has subtle influences on the wave signatures that we have not captured in this analysis.

This 3-yr periodic mode may prove important to predicting periods of high incidence of disease. The low periodicity of this mode facilitates prediction on annual timescales, which will be

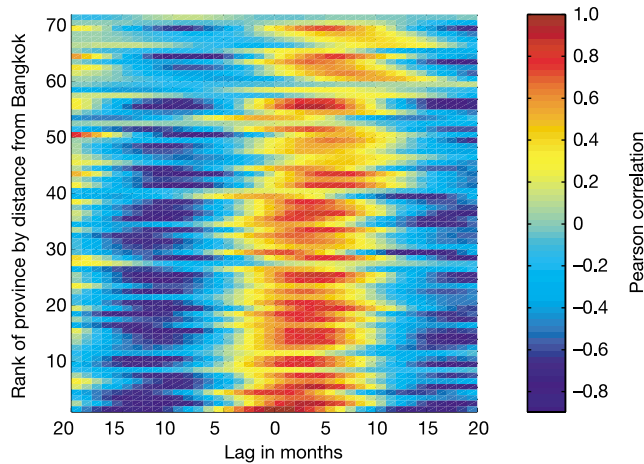


Figure 5 Cross-correlation coefficients between the 3-yr oscillatory mode of DHF incidence in Bangkok and the same mode of DHF incidence in the 71 other provinces of Thailand. Pearson correlation coefficients are depicted as colours for each monthly pairing, with red indicating strong positive correlation, and blue indicating strong negative correlation (see colour bar on right). CCFs are ordered by rank of distance from Bangkok.

useful in health system preparations. In addition, surveillance in Bangkok may prove useful for preparations in the surrounding regions, as epidemics are heralded by as much as ten months in some areas. Urban centres have been thought to be important in the genesis of dengue epidemics elsewhere in Asia²³. Our results suggest that high priority should be placed on surveillance systems in urban areas of Southeast Asia.

In this analysis, time series decomposition revealed a phenomenon that was not apparent in the raw incidence data or in other modes of variance in the incidence data. Although the isolation of particular modes of variance is a simplification of complex and dynamically interacting disease processes, these techniques can aid the formation of hypotheses and larger models. To our knowledge, this is the first application of EMD to the analysis of epidemiological data. We consider that the ability of the EMD to decompose nonstationary and nonlinear data makes it the most appropriate technique for this analysis. □

Methods

Data

Numbers of DHF cases have been collected by the Ministry of Health of Thailand since 1972. Cases are diagnosed on the basis of criteria of the World Health Organization (WHO). Serological confirmation is conducted where feasible, but logistically is impossible to do on every case report. Here we consider only the data that were available to us, that is, the monthly incidence of DHF in each of the 72 provinces of Thailand from 1983 to 1997. These data are available on the Johns Hopkins Center for Immunization Research website (<http://www.jhsph.edu/cir/dengue.html>). We are currently working to obtain incidence data from 1997 to the present.

EMD

The method of EMD decomposes a time series into IMFs by means of a sifting process⁴. The sifting process begins with the identification of local minima and maxima of a raw time series, $X(t)$. Two cubic splines are fit: one connecting the local maxima and one connecting the local minima. The time series of means of these two splines, $m_1(t)$, is calculated. The difference between the raw time series and the mean series, $X(t) - m_1(t)$, is designated $h_1(t)$. The series $h_1(t)$ is the first IMF of the data if it satisfies two admission criteria: the number of extrema and zero-crossings must not differ by more than 1, and the mean series between cubic splines connecting the extrema of $h_1(t)$ must be 0 at all times. If $h_1(t)$ does not satisfy these criteria, the algorithm is repeated using $h_1(t)$ as the raw series. The first IMF is subtracted from the raw data series, and the algorithm is repeated on this difference to identify subsequent IMFs of the data.

Log-transformations of incidence time series from each province normalized to have zero mean and unit standard deviation were decomposed²⁴. IMFs with an approximate period (measured by Fourier analysis) of 3–4 yr (the third IMF component for all but one province) were used in subsequent analyses. The Hilbert–Huang Transformation

Toolbox for Matlab (Princeton Satellite and NASA, 2001) was used for the EMD analysis.

Spatial synchrony, phase coherence and CCFs

Algorithms in the NCF library for R/S-plus (available at <http://asi23.ent.psu.edu/>) were used to estimate the spatial correlation functions and phase coherence functions. A detailed discussion of these techniques is given in refs 12, 18. Spatial synchrony of two DHF incidence time series was quantified as the Pearson correlation coefficient of those series¹⁷. Spatial correlation functions were estimated by the nonparametric spline covariance function¹⁸. A cubic B-spline of nine equivalent degrees of freedom was used in this estimation (the square root of the number of provinces was used as a guide¹⁸). We calculated C.I.s using 500 bootstrap iterations.

The phase coherence of two DHF incidence time series was calculated as the Pearson correlation coefficient of phase angles of each series¹². Again, the nonparametric spline covariance function was used to estimate the phase coherence function¹². We calculated CCFs using Pearson correlation coefficients. The lag at which each CCF was at its maximum absolute value (for lags between –12 and 12) was identified for each province.

Received 18 August; accepted 20 November 2003; doi:10.1038/nature02225.

- Gubler, D. J. Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* **11**, 480–496 (1998).
- Gubler, D. J. The global pandemic of dengue/dengue haemorrhagic fever: current status and prospects for the future. *Ann. Acad. Med. Singapore* **27**, 227–234 (1998).
- DeRoock, D., Deen, J. & Clemens, J. D. Policymakers' views on dengue fever/dengue haemorrhagic fever and the need for dengue vaccines in four southeast Asian countries. *Vaccine* **22**, 121–129 (2003).
- Huang, N. E. *et al.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A* **454**, 903–995 (1998).
- Ferguson, N., Anderson, R. & Gupta, S. The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proc. Natl Acad. Sci. USA* **96**, 790–794 (1999).
- Hay, S. I. *et al.* Etiology of interepidemic periods of mosquito-borne disease. *Proc. Natl Acad. Sci. USA* **97**, 9335–9339 (2000).
- Chareonsook, O., Foy, H. M., Teerarattul, A. & Silarug, N. Changing epidemiology of dengue hemorrhagic fever in Thailand. *Epidemiol. Infect.* **122**, 161–166 (1999).
- Watts, D. M., Burke, D. S., Harrison, B. A., Whitmore, R. E. & Nisalak, A. Effect of temperature on the vector efficiency of *Aedes aegypti* for dengue 2 virus. *Am. J. Trop. Med. Hyg.* **36**, 143–152 (1987).
- Tun-Lin, W., Burkot, T. R. & Kay, B. H. Effects of temperature and larval diet on development rates and survival of the dengue vector *Aedes aegypti* in north Queensland, Australia. *Med. Vet. Entomol.* **14**, 31–37 (2000).
- Gubler, D. J. & Rosen, L. Quantitative aspects of replication of dengue viruses in *Aedes albopictus* (Diptera: Culicidae) after oral and parenteral infection. *J. Med. Entomol.* **13**, 469–472 (1977).
- Rosen, L., Roseboom, L. E., Gubler, D. J., Lien, J. C. & Chaniotis, B. N. Comparative susceptibility of mosquito species and strains to oral and parenteral infection with dengue and Japanese encephalitis viruses. *Am. J. Trop. Med. Hyg.* **34**, 603–615 (1985).
- Grenfell, B. T., Bjornstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
- Bjornstad, O. N., Peltonen, M., Liebhold, A. M. & Baltensweiler, W. Waves of larch budmoth outbreaks in the European alps. *Science* **298**, 1020–1023 (2002).
- Bacon, P. J. *Population Dynamics of Rabies in Wildlife* (Academic, London, 1985).
- Sherratt, J. A. Periodic travelling waves in cyclic predator–prey systems. *Ecol. Lett.* **4**, 30–37 (2001).
- Mollison, D. Modeling biological invasions—chance, explanation, prediction. *Phil. Trans. R. Soc. Lond. B* **314**, 675–693 (1986).
- Bjornstad, O. N., Ims, R. A. & Lambin, X. Spatial population dynamics: analyzing patterns and processes of population synchrony. *Trends Ecol. Evol.* **14**, 427–432 (1999).
- Bjornstad, O. N. & Falck, W. Nonparametric spatial covariance functions: estimation and testing. *Environ. Ecol. Stat.* **8**, 53–70 (2001).
- Lambin, X., Elston, D. A., Petty, S. J. & MacKinnon, J. L. Spatial asynchrony and periodic travelling waves in cyclic populations of field voles. *Proc. R. Soc. Lond. B* **265**, 1491–1496 (1998).
- de Roos, A. M., McCauley, E. & Wilson, W. G. Pattern formation and the spatial scale of interaction between predators and their prey. *Theor. Popul. Biol.* **53**, 108–130 (1998).
- Coulson, T. *et al.* Age, sex, density, winter weather, and population crashes in Soay sheep. *Science* **292**, 1528–1531 (2001).
- Nisalak, A. *et al.* Serotype-specific dengue virus circulation and dengue disease in Bangkok, Thailand from 1973 to 1999. *Am. J. Trop. Med. Hyg.* **68**, 191–202 (2003).
- Gubler, D. J. *et al.* Virological surveillance for dengue hemorrhagic-fever in Indonesia using the mosquito inoculation technique. *Bull. WHO* **57**, 931–936 (1979).
- Anderson, R. M., Grenfell, B. T. & May, R. M. Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: time series analysis. *J. Hyg. (Lond.)* **93**, 587–608 (1984).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank O. Bjornstad for making R code available. This work was supported by grants from the National Oceanic and Atmospheric Administration's Joint Program on Climate Variability and Human Health (a consortium including the EPA, NASA, NSF and EPRI), and the Bill and Melinda Gates Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.S.B. (dburke@jhsph.edu).

A primitive Y chromosome in papaya marks incipient sex chromosome evolution

Zhiyong Liu¹, Paul H. Moore², Hao Ma^{1,3}, Christine M. Ackerman¹, Makandar Ragiba¹, Qingyi Yu^{1,3}, Heather M. Pearl¹, Minna S. Kim¹, Joseph W. Charlton¹, John I. Stiles⁴, Francis T. Zee², Andrew H. Paterson⁵ & Ray Ming¹

¹Hawaii Agriculture Research Center, Aiea, Hawaii 96701, USA

²USDA-ARS, Pacific Basin Agricultural Research Center, Hilo, Hawaii 96720, USA

³Department of Molecular Biosciences and Bioengineering, University of Hawaii, Honolulu, Hawaii 96822, USA

⁴Integrated Coffee Technologies Incorporated, Waialua, Hawaii 96791, USA

⁵Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia 30602, USA

Many diverse systems for sex determination have evolved in plants and animals^{1–3}. One involves physically distinct (heteromorphic) sex chromosomes (X and Y, or Z and W) that are homozygous in one sex (usually female) and heterozygous in the other (usually male). Sex chromosome evolution is thought to involve suppression of recombination around the sex determination genes, rendering permanently heterozygous a chromosomal region that may then accumulate deleterious recessive mutations by Muller's ratchet, and fix deleterious mutations by hitchhiking as nearby favourable mutations are selected on the Y chromosome^{4,5}. Over time, these processes may cause the Y chromosome to degenerate and to diverge from the X chromosome over much of its length; for example, only 5% of the human Y chromosome still shows X–Y recombination⁶. Here we show that papaya contains a primitive Y chromosome, with a male-specific region that accounts for only about 10% of the chromosome but has undergone severe recombination suppression and DNA sequence degeneration. This finding provides direct evidence for the origin of sex chromosomes from autosomes.

Papaya (*Carica papaya* L., $2n = 18$), a polygamous angiosperm with male, female and hermaphroditic forms, offers several advantages for genetic and evolutionary studies. These include a small genome of 372 megabases (Mb), a short generation time of 9–15 months, numerous flower types, a unique evolutionary process in female flowers (see Supplementary Information), an intriguing system of sex determination and an established transformation system. Sex determination in papaya has been a frequent subject of genetic analyses^{7–10} because it is directly related to efficient commercial fruit production.

Hofmeyr⁷ and Storey⁸ each concluded that sex in papaya is determined by a single gene with three alleles: M, male; M^h, hermaphrodite; and m, female. Females (mm) are homozygous recessive. Males (Mm) and hermaphrodites (M^hm) are enforced sex heterozygotes, with typically 25% of seeds nonviable in fruit, because every combination of dominant alleles (MM, MM^h or M^hM^h) is embryonic lethal. Hofmeyr⁹ later suggested that M and M^h represent genetically inactive regions of the sex chromosomes

from which vital genes are missing. In Storey's¹⁰ final hypothesis, the sex-determining region includes a complex of genes regulating stamen and carpel development, a lethal factor and a recombination-suppressing factor. Because males and hermaphrodites are heterogametic, whereas females are homogametic, sex determination in papaya has been considered by some to be of the XY chromosome type even though heteromorphic chromosomes have not been found^{1,11}.

The papaya sex locus has been genetically mapped to linkage group 1 (LG 1; ref. 12). In a high-density genetic map containing 1,501 amplified fragment length polymorphism (AFLP) and morphological markers from 54 F₂ plants, a total of 225 markers (that is, 15% of all markers mapped on the genome and 66% of the 342 markers on LG 1) co-segregate with the sex locus¹³. The remaining 117 AFLP markers on LG 1 detect regions of active recombination spanning 289.7 cM. These data reveal an extremely high level of DNA polymorphism in the genomic region immediately surrounding the sex locus.

We fine-mapped the sex determination gene (Table 1 and Supplementary Fig. 1), using 4,380 informative chromosomes (two each from 2,190 female and hermaphrodite plants from three F₂ and one F₃ populations), and two sequence-characterized amplified region (SCAR) markers (W11 and T12)¹⁴, three cloned sex-linked AFLP markers (cpm10, cpm31 and cpm54) and one cloned bacterial artificial chromosome (BAC) end (cpbe55). No recombinants were detected. The finding that a region so recombinationally small contains such a large percentage of polymorphic markers indicates that the two homologues are highly differentiated in this region. This is consistent with the classical notion that an early stage of sex chromosome evolution involves suppression of recombination around the sex determination locus, leading to gradual degeneration of the Y chromosome.

We physically mapped the non-recombining region by using a 13.7 × BAC library¹⁵. The sex-linked SCAR marker W11 identified four positive BACs. A BAC contig assembled by chromosome walking spanned 990 kilobases (kb). Forty-two cpm markers and three previously available sex co-segregating markers, T12, PSDM and Nafp (see Supplementary Information), hybridized to additional BACs. Among these 42 cpm markers, 24 (57%) could be placed on BAC contig maps, 3 identified individual contigs (data not shown), and 15 contained repetitive sequences (and thus could not be mapped). We cloned (iteratively) 92 BAC ends from contig-terminal BACs and used them to close the gaps. All relevant BAC clones were fingerprinted and mapped to verify the hybridization-based physical map. The 2.5-Mb physical map includes two major and three minor contigs, containing 4 SCAR, 82 cpbe and 24 cpm loci (Fig. 1).

Although gaps still remain in the physical map, at least 57% of the cpm markers are located in a small (2.5-Mb) region of the sex-determining chromosome. On the basis of the papaya's genome size and the cytological observation that its chromosomes are similar in size¹⁶, each papaya chromosome is estimated to be at about 41 Mb. With a mapping population of 54 plants¹³ and the use of dominant markers (that is, only one gamete is genetically informative), the absence of recombinants suggests that two loci must be less than about 4 cM apart with 95% confidence¹⁷, which delimits the maximum size of the male-specific region (MSY). The lack of recombinants in 2,190 plants suggests that the 225 sex

Table 1 Results of fine-mapping with SCAR markers in the MSY region

Population	Progeny	Hermaphrodite	Female	SCAR markers	Recombinant
Kapoho × SunUp	F ₂	335	150	W11	0
Kapoho × SunUp	F ₂	335	156	W11,T12, cpbe55	0
Kapoho × SunUp	F ₃	481	274	W11,T12, cpbe55	0
Kapoho × Saipan Red	F ₂	175	49	cpm31, cpm10	0
AU9 × SunUp	F ₂	170	65	W11,T12, cpm54	0
Total	2,190	1,496	694		0

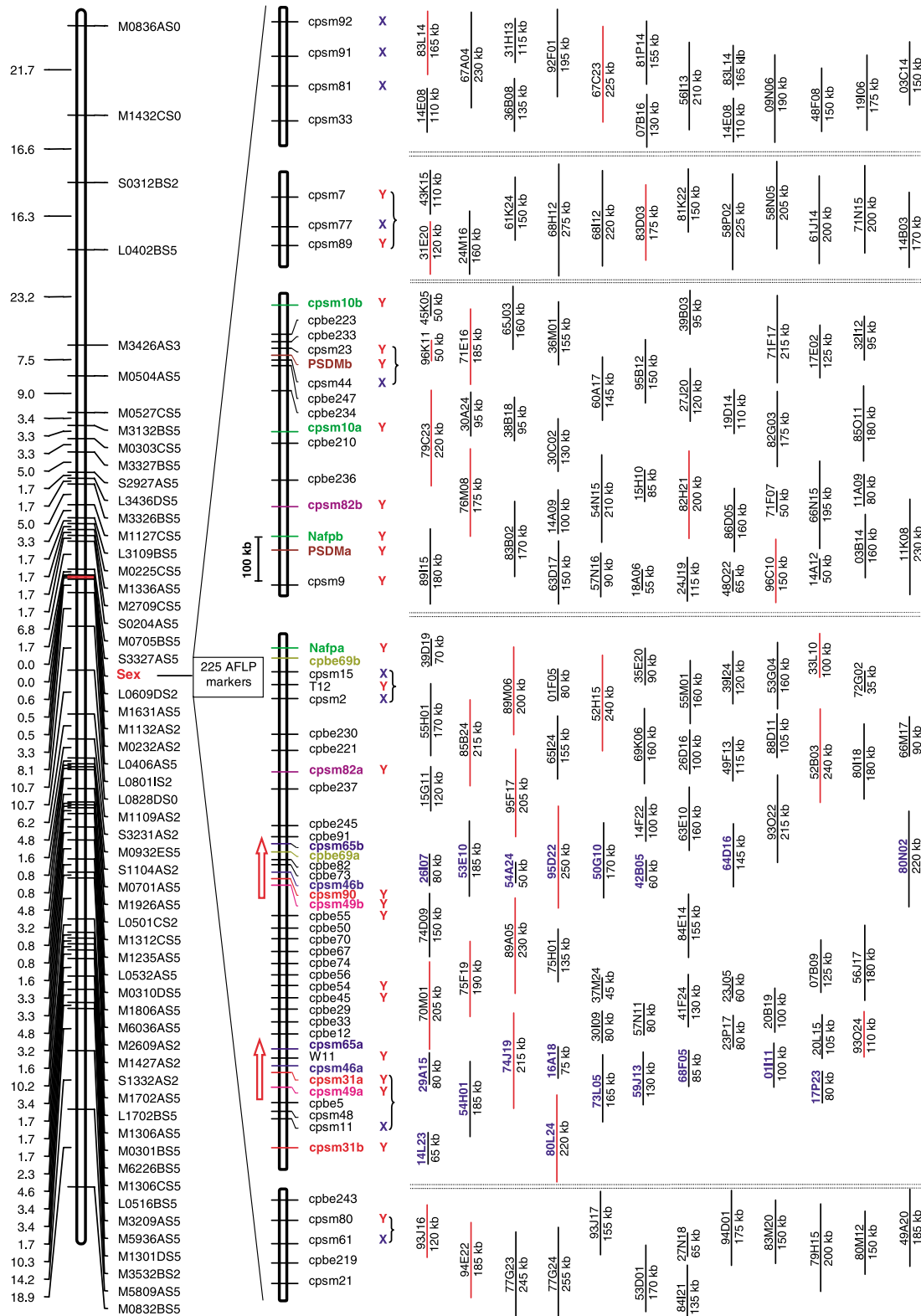


Figure 1 AFLP map of papaya LG 1 and physical maps of BAC contigs in the MSY region. Left, AFLP map of LG 1. Right, physical maps of two long and three short BAC contigs in the MSY region. Arrows indicate a large tandem duplication; marker names shown in the same colour indicate additional smaller duplications. The letters X (blue) and Y (red)

indicate whether the sequence is X-like or Y-like. The X- and Y-like DNA sequences located within one BAC are delineated by a bracket. BAC clones used for sequencing from the MSY are shown in red. BAC clones tested in Fig. 2c are marked in blue.

co-segregating markers are located in a region that is much less than 4 cM. The remaining 117 markers on LG 1 span 289.7 cM, however, indicating that much of this chromosome is still recombining with its homologue¹³. Among the markers used for fine-mapping, W11 and T12 are about 900 kb apart, as estimated by adding the insert sizes of non-overlapping BAC clones. On the basis of the total genetic map of 3,294 cM and the genome size of 372 Mb, this is equivalent to 8 cM in other genomic regions with average recombination rates. Thus, the 1,481 plants informative in fine-mapping with both W11 and T12 would have been expected to contain 237 recombinant plants, but none was found.

A mosaic arrangement of conserved (X-like) and diverged (Y-like) sequences (Fig. 1) shows that degradation of the Y chromosome is distributed across the non-recombining region. Out of 26 test sequences designed from the physically mapped non-recombining region in male, female and hermaphrodite plants (see Methods), 17 were Y-like (Fig. 1, 'Y') and present in hermaphrodite and male but not female, whereas 9 were X-like (Fig. 1, 'X') and present in all three sex types. Six X-like sequences were on BACs that also contained Y-like sequences (Fig. 1, brackets).

Duplicated sequences are common in the non-recombining region. Out of 109 DNA markers on the physical map, 9 identified duplications, including 5 (31%) Y-like test sequences. A tandem duplication identified by cpsm31, cpsm46, cpsm49, cpsm65 and cpsm90 spanned more than 100 kb, with the repeats located 500 kb apart (Fig. 1). Nucleotide insertions, deletions and substitutions

were observed in all duplicated DNA sequences such as cpsm31 and cpsm90 (Fig. 2 and Supplementary Fig. 2).

Random subclone sequences (628 reads, totalling 513 kb) from 25 non-redundant BACs of the non-recombining region were analysed to determine the nature of the sequences and to estimate the gene density. We found that 48 reads from 44 subclones matched functional genes, suggesting an average density of 8.6 genes per 100 kb. Using RepeatMasker (<http://repeatmasker.genome.washington.edu>) to search against the *Arabidopsis thaliana* repeat database, we identified 95 long terminal repeat (LTR) retroelements (including 82 Gypsy-like and 13 Copia-like elements), with an average of 18.5 LTRs per 100 kb. Forty short inverted repeats were identified (7.8 per 100 kb). In comparison to a genome-wide sample of papaya DNA based on 684 reads totalling 517 kb (GenBank Accessions CG026197 to CG026996), the MSY region showed 37.7% lower gene density, 27.6% higher retroelement density, and 188.9% higher inverted repeat density.

The severe suppression of recombination and extensive divergence between homologues in the region containing the papaya sex-determining genes indicates that this is an incipient sex chromosome. On the basis of the size of the present contig map (2.5 Mb) and the 57% of cpsm markers that have been accounted for, the physical size of the MSY is estimated at 4–5 Mb or 10% of papaya's primitive Y chromosome.

The discovery of an incipient Y chromosome in papaya, of which 10% is a non-recombining, rapidly evolving, sex-determining region flanked by normal autosome-like regions that comprise the remaining 90% of the chromosome (Fig. 3), provides direct evidence for the theory that sex chromosomes evolve from autosomes³. It also supports the hypothesis that sex chromosomes arise from suppression of recombination around the sex-determining locus^{1,3,18}. The papaya MSY region may resemble the ancestor of the human Y chromosome as it existed about 240–320 million years ago, when recombination suppression is postulated to have been limited to only a small portion of the chromosome¹⁹. Analysis of the pseudoautosomal boundary in mammals indicates that sex chromosome differentiation in humans probably started near male-determining loci by suppression of homologous recombination²⁰.

The incipient sex chromosomes of papaya may yield insights about earlier stages of sex chromosome evolution. The small physical size of the MSY region and the mosaic arrangement of sequence degradation indicate a recent origin of the papaya sex chromosomes (see Supplementary Information). By contrast, in the better-studied X and Y chromosomes of the plant *Silene latifolia*, 90% of the Y chromosome, albeit mostly euchromatic DNA, seems to be degenerated^{1,21,22}. Although heteromorphic sex chromosomes have not been detected in papaya, a high frequency of precocious separation of one chromosome pair during meiosis has been reported in the anthers of male and hermaphrodite plants but not in the pistils of female plants (ref. 24 and Supplementary Fig. 3), which is compatible with relaxed pairing between primitive Y and X chromosomes owing to the existence of a non-pairing MSY region.

The lethal effect of homozygous dominant sex-determining alleles in papaya^{7–10} provides further evidence of the degeneration of the MSY region. Because papaya seeds with the homozygous Y chromosomes are aborted 25–50 d after pollination²³, genes that are

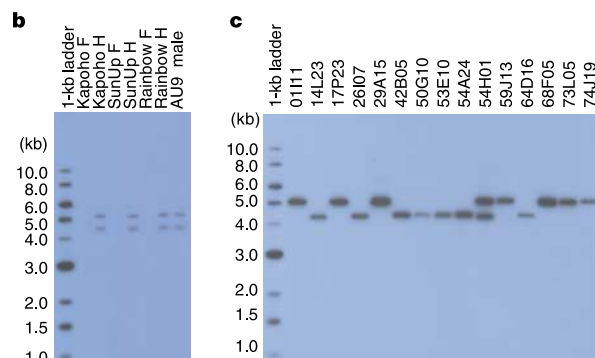
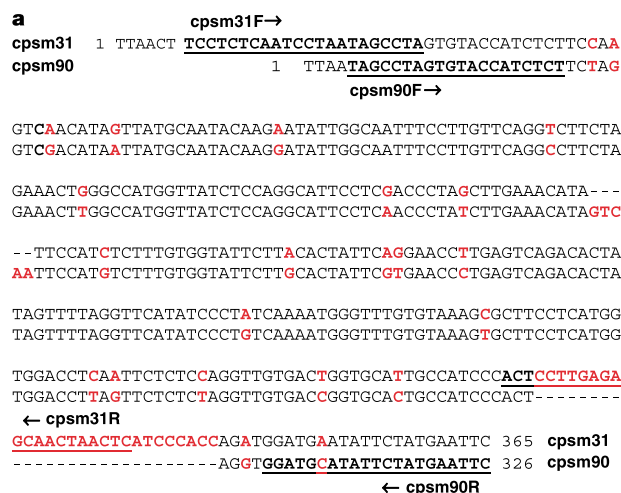


Figure 2 Characterization of duplicated markers cpsm31 and cpsm90 in the MSY region. **a**, DNA sequence comparison of cpsm31 and cpsm90. PCR primer sequences are underlined. Nucleotide insertion, deletion and transition or transversion are indicated in red. **b**, Southern hybridization pattern of cpsm31 on *Hind*III-digested genomic DNA of different sex types. Two cpsm31 loci are present in hermaphrodites (H) and males, but are absent from females (F). **c**, Southern hybridization of cpsm31 on *Hind*III-digested BAC DNA from the tandemly duplicated region of MSY.

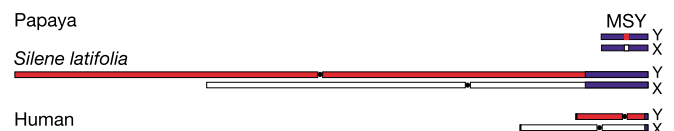


Figure 3 Comparative organization of Y chromosomes in papaya (estimated as 41 Mb), *S. latifolia* (estimated as 570 Mb) and human (66 Mb). The extent of MSY regions (red) and recombinationally active regions (blue) is shown. Filled circles represent the centromeres. The location of the centromere on the papaya Y chromosome remains unknown.

vital for papaya embryo development may once have existed in what is now the MSY region. Future detailed comparisons of the MSY and its homologue, as well as related chromosomes from other taxa that do not have discrete sexes, may help to identify these genes.

DNA sequence duplication and repetition in the MSY regions of papaya, *Drosophila* and humans may reflect common processes associated with evolution of the sex chromosomes in plants and animals. Complete sequencing of the euchromatic DNA on the human Y chromosome has identified X-transposed, X-degenerated and ampliconic sequences⁶. X-degenerated sequences seem to be common in the papaya MSY, because 65% of the test sequences were diverged between the primitive X and Y chromosomes and about a third of the MSY is degenerated on the basis of gene density. Ampliconic sequences are probably present on the papaya primitive Y chromosome, because abundant inverted repeats were found and 38% of random DNA markers in the MSY detected repetitive sequences. Sequence duplication and retroelement accumulation may be driving the degeneration of the papaya MSY region. X- and Y-linked homologues in *Silene* also show deletion and accumulation of repetitive sequences, as well as low variability in a Y-linked gene²². The liverwort Y chromosome is a mosaic structure of repeats and subrepeats²⁵. In *Drosophila miranda*, genes show various signs of degeneration on the non-recombining neo-Y, and repetitive DNA is accumulated²⁶.

We found that hermaphrodite and male papaya plants share identical DNA sequences in most parts of the MSY region (Supplementary Fig. 2 and Table S1). These two sex types seem to share a haplotype for this region that differs from that of females and is recently derived from a common ancestral chromosome. This observation suggests that the hermaphrodites with an MSY region are already genetically different from the ancestral hermaphrodites. The findings in papaya strongly support the hypothesis that Y chromosome evolution involved at least two steps. First, a single male-sterility mutation led to femaleness and to selection for absence of recombination; second, the Y chromosome evolved as hermaphrodites converted to males by a second mutation or mutations led to female sterility¹⁸. Thus, gynodioecy might be an intermediate step to dioecy in the family Caricaceae²⁷.

Papaya provides an extraordinary opportunity to elucidate the evolutionary process of dioecy because monoecious, dioecious and perfect flowers all exist among its close relatives. Only two species, *C. papaya* and *Vasconcellea cundinamaricensis*, are polygamous with all three sex types; *V. monoica* is monoecious; and the other 32 species in the family Caricaceae are dioecious. DNA sequencing will allow comparative analyses of the MSY region on the primitive Y chromosome of monoecious and dioecious species of Caricaceae to shed light on the evolution of sex chromosomes and dioecy.

Although higher plants can regulate sex by either autosomal genes or sex chromosomes²¹, few flowering plants have evolved heteromorphic sex chromosomes analogous to those of many animals. Despite the recent origin of plant sex chromosomes, they share several characteristics with those in animals, such as prevalence of male heterogamy, degeneration of the Y chromosome, and lack of recombination between X and Y chromosomes. These parallel properties indicate that many similar, if not identical, evolutionary forces may drive sex chromosome evolution in plants and animals. The primitive sex chromosomes in papaya provide a unique opportunity to test hypotheses and theories about sex chromosome evolution near the inception of this process¹⁸. □

Methods

Plant materials

Female and hermaphrodite plants of the Hawaiian Solo type papaya cultivars Kapoho and SunUp and their F₁ progeny Rainbow, as well as male plants of the Australian variety AU9, were used for polymerase chain reaction (PCR) and Southern hybridization analyses. We used three F₂ populations derived from single self-pollinated hermaphrodite F₁ plants from each of the three crosses: Kapoho × SunUp, Kapoho × Saipan Red and

AU9 × SunUp. The F₃ population of Kapoho × SunUp was from self-pollinated hermaphrodite F₂ plants.

DNA markers used for physical mapping

Three kinds of sex-linked DNA marker were used to screen the papaya BAC library and to construct the BAC contigs: SCAR markers from previous research¹⁴, *Carica papaya* BAC ends (cpbe) cloned by a PCR approach²⁸, and cloned sex-linked AFLP markers (cpsm) identified from a papaya high-density map¹³. We excised the sex-linked AFLP fragments from manual sequencing gels after silver staining and cloned them into a pCRII-TOPO vector (Invitrogen). DNA sequencing was done on a DNA sequencer (Li-Cor) by using infrared dye-labelled primers. PCR primers were designed for each cpsm and were used to amplify genomic DNA from females and hermaphrodites of the cultivars Kapoho, SunUp and Rainbow, and from a male of AU9.

Contig map construction

We screened our papaya BAC library with sex-linked DNA markers to identify putative BAC clones, which were confirmed by Southern hybridization. We then constructed preliminary BAC contigs based on Southern hybridization of all BAC ends to marker-positive BAC clones. The ends of the terminal BACs from each contig were used to re-screen the BAC library to extend the contigs. In parallel, all BACs in the MSY region were digested with *HindIII* and fingerprinted by separation on agarose gels. Gel images were analysed with Image3 software (Sanger Institute) and Fingerprinting Contigs (FPC) for confirmation of the assembled contigs.

Development of *Carica papaya* sex-linked markers (cpsm)

The 225 sex co-segregating AFLP markers were surveyed to identify 85 markers that were easily resolved from neighbouring bands and were within the optimum size of 200–500 base pairs to be used as probes for physical mapping. These markers were excised from AFLP gels to develop *Carica papaya* sex-linked markers (cpsm). Sixty-three isolated cpsm fragments were reamplified with the original AFLP primers and confirmed by running the PCR products and the original AFLP products in a sequencing gel.

Test sequences in the non-recombining region

Twenty-six sequences from the non-recombining region were tested to see whether they are conserved or divergent. Nineteen test sequences were derived from the 24 cpsm markers used for physical mapping by sequencing the cpsm markers and by designing pairs of primers to amplify female, hermaphrodite and male genomic DNA. The other seven test sequences were derived from three BAC ends (cpbe45, cpbe54 and cpbe55) and from the original four RAPD-derived SCAR markers (W11, T12, PSDM and Nafp).

Subcloning and sequencing of BAC clones in the MSY

To examine the nature of DNA sequences in the MSY of the papaya primitive Y chromosome, we selected non-redundant BAC clones for sequencing. Twenty-five BACs were digested with *HindIII* and subcloned into the pPCR-Script Amp vector (Stratagene). Forty subclones were grown out for each BAC and their DNA was isolated. After *HindIII* digestion to release the inserts, 10–15 subclones of each BAC with different insert sizes were sequenced from both ends.

Gene density estimation

We randomly sequenced 628 reads, which were sampled roughly equally from 25 non-redundant BAC subclones totalling 513 kb. These were used for a blast search of the National Center for Biotechnology Information (NCBI) nucleotide sequence databases nr and est_others using BLASTN (<http://www.ncbi.nlm.nih.gov/BLAST/>). Functional gene sequence matches were accepted with *E*-value thresholds of less than e^{-10} . Eight hundred papaya genome survey sequences (C. Town, personal communication) totalling 605 kb were also used in a BLASTN search. We eliminated 14 mitochondrial DNA and 102 chloroplast DNA sequences, and used 684 sequences totalling 517 kb for estimating papaya genome-wide gene density.

Estimation of the sex chromosome size in papaya and *S. latifolia*

The size of the papaya sex chromosomes is estimated on the basis that the nine pairs of chromosomes have similar sizes by cytological assessment¹⁶ and that the genome size is 372 Mb. The estimated size of the X (400 Mb) and Y (570 Mb) chromosomes in *S. latifolia* is based on the physical measurement of each chromosome from its karyotype and the female and male genome size of 5,529 Mb/2C and 5,645 Mb/2C (ref. 29), respectively. The proportion of X and Y chromosome arms in *S. latifolia* has been described³⁰. There are no data on the length of the pseudoautosomal region (PAR) in *S. latifolia*, and this is a rough estimate (and thus not drawn to scale in Fig. 3).

Received 4 August; accepted 13 November 2003; doi:10.1038/nature02228.

- Westergaard, M. The mechanism of sex determination in flowering plants. *Adv. Genet.* **9**, 217–281 (1958).
- White, M. J. D. In *Insect Reproduction* (eds Leather, S. R. & Hardie, J.) 57–94 (Cambridge Univ. Press, Cambridge, 1973).
- Ohno, S. *Sex Chromosome and Sex-linked Genes* (Springer, Berlin, 1967).
- Muller, M. The relation of recombination to mutational advance. *Mutat. Res.* **1**, 2–9 (1964).
- Rice, W. R. Genetic hitch-hiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**, 161–167 (1987).
- Skaletsky, H. et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
- Hofmeyr, J. D. J. Genetical studies of *Carica papaya* L. *S. Afr. J. Agric. Sci. Bull.* **187**, 64 (1938).
- Storey, W. B. Segregation of sex types in *Solo papaya* and their application to the selection of seed. *Am.*

- Soc. Hortic. Sci. Proc. 35, 83–85 (1938).
9. Hofmeyr, J. D. J. Some genetic and breeding aspects of *Carica papaya*. *Agronomia Tropical* 17, 345–351 (1967).
 10. Storey, W. B. In *The Evolution of Crop Plants* (ed. Simmonds, N. W.) 21–24 (Longman, London, 1976).
 11. Horovitz, S. & Jimenez, H. Cruzamientos interespecificos e intergenericos en Caricaceas y sus implicaciones fitotecnias. *Agronomia Tropical* 17, 353–359 (1967).
 12. Sondur, S. N., Manshardt, R. M. & Stiles, J. I. A genetic linkage map of papaya based on randomly amplified polymorphic DNA markers. *Theor. Appl. Genet.* 93, 547–553 (1996).
 13. Ma, H. et al. High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. *Genetics* (in the press).
 14. Deputy, J. C. et al. Molecular marker for sex determination in papaya (*Carica papaya* L.). *Theor. Appl. Genet.* 106, 107–111 (2002).
 15. Ming, R. et al. Construction and characterization of a papaya BAC library as a foundation for molecular dissection of a tree-fruit genome. *Theor. Appl. Genet.* 102, 892–899 (2001).
 16. Storey, W. B. The botany and sex relations of the papaya. *Hawaii Agric. Exp. Station Bull.* 87, 5–22 (1941).
 17. Tanksley, S. D., Miller, J. C., Paterson, A. H. & Bernatzky, R. *Proc. 18th Stadler Genet. Symp.* 157–173 (Plenum, New York, 1988).
 18. Charlesworth, B. The evolution of sex chromosomes. *Science* 251, 1030–1033 (1991).
 19. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* 286, 964–967 (1999).
 20. Iwase, M. et al. The *amelogenin* loci span an ancient pseudoautosomal boundary in diverse mammalian species. *Proc. Natl Acad. Sci. USA* 100, 5258–5263 (2003).
 21. Charlesworth, D. Plant sex determination and sex chromosomes. *Heredity* 88, 94–101 (2002).
 22. Filatov, D. A., Moneger, F., Negritiu, I. & Charlesworth, D. Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* 404, 388–390 (2000).
 23. Chiu, C. T. *Study on Sex Inheritance and Horticultural Characteristics of Hermaphrodite Papaya* Thesis, Nat. Pingtung Univ. Sci. Technol., Taiwan (2000).
 24. Storey, W. B. Genetics of the papaya. *J. Hered.* 44, 70–78 (1953).
 25. Okada, S. et al. The Y chromosome in the liverwort *Marchantia polymorpha* has accumulated unique repeat sequences harboring a male-specific gene. *Proc. Natl Acad. Sci. USA* 98, 9454–9459 (2001).
 26. Bachtrog, D. Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in *Drosophila*. *Nature Genet.* 34, 215–219 (2003).
 27. Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* 112, 975–997 (1978).
 28. Yang, Z. N. & Mirkov, T. E. Isolation of large terminal sequences of BAC inserts based on double restriction enzyme digestion followed by anchored PCR. *Genome* 43, 412–415 (2000).
 29. Siroky, J., Lysak, M. A., Dolezel, J., Kejnovsky, E. & Vyskot, B. Heterogeneity of rDNA distribution and genome size in *Silene* spp. *Chromosome Res.* 9, 387–393 (2001).
 30. Lengerova, M., Moore, R. C., Grant, S. R. & Vyskot, B. The sex chromosomes of *Silene latifolia* revisited and revised. *Genetics* 165, 935–938 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Charlesworth for comments on the manuscript; R. Perl-Treves for discussions; S. Ancheta, G. Asmus and L. Poland for technical assistance; R. Manshardt for providing an F₂ population for fine-mapping; and H. Albert, M. Moore, R. Osgood, B. Vyskot and S. Whalen for reviewing the manuscript. This work was supported by a United States Department of Agriculture Agricultural Research Service (USDA-ARS) Cooperative Agreement with the Hawaii Agriculture Research Center, and a subaward to R. M. and A.H.P. to produce and to characterize the BAC library.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M. (rmimg@harc-hspa.com). The sequences are deposited in GenBank under accession codes CG680451–CG681045 and AY428881–AY428946.

Sleep inspires insight

Ullrich Wagner¹, Steffen Gais¹, Hilde Haider², Rolf Verleger³ & Jan Born¹

¹Department of Neuroendocrinology, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

²Institute of Psychology, University of Cologne, Gronewaldstrasse 2, 50931 Cologne, Germany

³Department of Neurology, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

Insight denotes a mental restructuring that leads to a sudden gain of explicit knowledge allowing qualitatively changed behaviour^{1,2}. Anecdotal reports on scientific discovery suggest that pivotal insights can be gained through sleep³. Sleep consolidates recent memories^{4–6} and, concomitantly, could allow insight by

changing their representational structure. Here we show a facilitating role of sleep in a process of insight. Subjects performed a cognitive task requiring the learning of stimulus–response sequences, in which they improved gradually by increasing response speed across task blocks. However, they could also improve abruptly after gaining insight into a hidden abstract rule underlying all sequences. Initial training establishing a task representation was followed by 8 h of nocturnal sleep, nocturnal wakefulness, or daytime wakefulness. At subsequent retesting, more than twice as many subjects gained insight into the hidden rule after sleep as after wakefulness, regardless of time of day. Sleep did not enhance insight in the absence of initial training. A characteristic antecedent of sleep-related insight was revealed in a slowing of reaction times across sleep. We conclude that sleep, by restructuring new memory representations, facilitates extraction of explicit knowledge and insightful behaviour.

The idea that sleep can trigger the gain of insight is associated with the names of famous scientific discoverers³. For example, Nobel prize winner Loewi reported that he woke up with the essential idea for an experimental confirmation of his theory of chemical neurotransmission. Mendeleyev, who laid out the periodic table of chemical elements, reported that his understanding of the critical rule underlying it emerged out of a dream following unsuccessful puzzling with the symbols of the elements. Recent studies in animals and humans provided evidence for the concept that neuronal representations of task stimuli and responses acquired during wakefulness become reactivated during subsequent sleep^{7–10}. This reprocessing of representations is considered to underlie the consolidating effect of sleep on memory^{4,11–15}, but could also be accompanied by restructuring these representations in memory to enable insight. Here, we tested whether sleep influences task representations in memory such that the gain of insight is facilitated. This was expressed in the extraction of explicit knowledge of a hidden abstract rule in stimulus–response sequences learned previously under implicit conditions.

To grasp the inherently unpredictable phenomenon of insight experimentally, we used a modified version of the Number Reduction Task (NRT; Fig. 1a) originally developed by Thurstone and Thurstone^{16–18}. Based on continuous monitoring of subjects' behavioural responses, the task allows the exact determination of the time point when insight occurs, that is, when explicit knowledge of a hidden abstract rule is gained, leading to an abrupt, qualitative shift in responding. On each trial of the task, subjects were asked to transform a given string of eight digits into a new string through a stepwise digit-by-digit application of two simple rules to reach a specific digit indicating the final solution to this string. With increasing practice, the subject's responses become gradually faster on this task. Most important, however, a hidden rule was implemented in the digit strings, which was not mentioned to the subjects and was therefore initially processed at an implicit level without awareness. The time point when a subject gained insight into this rule could be determined precisely because at this time he/she would begin to cut short sequential responding to confirm the final solution in advance. All subjects were first trained on three task blocks to induce mental representations of the task that still remained implicit with regard to the hidden rule during this period. The training period was then followed by an 8-h interval of (1) nocturnal sleep, (2) nocturnal wakefulness, or (3) daytime wakefulness (Fig. 1b). Subsequently, subjects were retested on ten blocks.

Sleep more than doubled the probability of gaining insight into the hidden rule compared to wakefulness. In the sleep group, thirteen out of 22 subjects (59.1%) gained insight at retesting, compared to five subjects (22.7%) in either wake group ($\chi^2 = 8.54$, degrees of freedom (d.f.) = 2, $P = 0.014$; Fig. 2). For subjects gaining insight, the time point of its occurrence (number of blocks after beginning of retesting) did not differ significantly between groups (sleep, 4.5 ± 0.8 (mean $\pm$ s.e.m.); wake-night, 6.8 ± 1.5 ;

wake-day, 6.0 ± 1.0 ; $P > 0.28$). The two nocturnal and daytime wake conditions excluded the possibility that inferior performance after wake intervals resulted from sleep deprivation or variations in circadian rhythm. Subjective ratings obtained at the beginning of retesting revealed that, as expected, subjects in the sleep condition were less tired compared to subjects tested following a wakeful night (five-point scale: 2.95 ± 0.22 versus 3.64 ± 0.25 , $P < 0.05$), but not compared to those tested after a period of daytime wakefulness (2.41 ± 0.22 , $P > 0.11$). This pattern excludes that the inferior performance in the wake conditions resulted from unspecific effects of tiredness.

To ensure that the influence of sleep was on memory representations established during training before sleep rather than a proactive influence on performance at retesting, in supplementary experiments subjects were tested on 13 continuous blocks either in the morning after sleep (7:00 h) or in the evening after daytime wakefulness (19:00 h), with only ten blocks included in the analysis for the direct comparison with the retest situation of the main experiment. Since, unlike in the main experiment, subjects had no

training before these test periods, off-line restructuring of initially acquired task representations could not occur during the periods of sleep or wakefulness, respectively, preceding task performance. Five out of 20 subjects (25.0%) gained insight in each of these two conditions, a level nearly identical to that of the wake conditions of the main experiment and substantially lower than in the sleep condition of the main experiment ($\chi^2 = 7.07$, d.f. = 2, $P = 0.029$; Fig. 2; the same result, $P < 0.01$, was obtained in this comparison when, as a control for overall task practice, the last rather than the first 10 of the 13 blocks were analysed). Thus, compared to subjects lacking initial training, subjects of the main experiment benefited from the 8-h post-training interval only if it was spent asleep. Critically, the sleep group of the main experiment also performed substantially better than subjects tested in the early morning after sleep without having established a task representation before sleep, ruling out an unspecific proactive influence of sleep on subsequent task performance.

Insightful behaviour can announce itself by characteristic antecedents^{19,20}. Our task allowed us to identify sleep-related behavioural antecedents of insight in characteristic changes of reaction time patterns across the sleep and wake intervals in the main experiment. Following previous work with the NRT¹⁸, we analysed the subjects' seven reaction times of each response string separately for the initial digit (response 1), the following three digits underdetermined by the task structure (responses 2–4), and the last three digits (responses 5–7), which were fully determined because they were always mirroring responses 2–4 (Fig. 1a). We identified changes in the three response types from the last block of initial training to the first block of retesting separately for 'solvers' (subjects who later gained insight into the hidden rule) and 'nonsolvers' (subjects not gaining insight). Overall, reaction times decreased across both sleep and wakefulness. However, this decrease differed saliently between solvers and nonsolvers in the sleep and wake conditions (the data of the two wake conditions did not differ and were combined in this analysis). Notably, in the sleep condition the solvers showed only marginal speeding of reaction times across sleep as compared to the profound decrease in the nonsolvers' reaction times (overall speeding 29.4 ± 27.2 ms in solvers versus 206.4 ± 35.1 ms in nonsolvers; F -test variance ratio ($F_{1,20}$) = 16.35, $P < 0.001$, for solver/nonsolver main effect; Fig. 3a). This differential influence of sleep was most obvious for the first response, which in solvers even slowed down across sleep

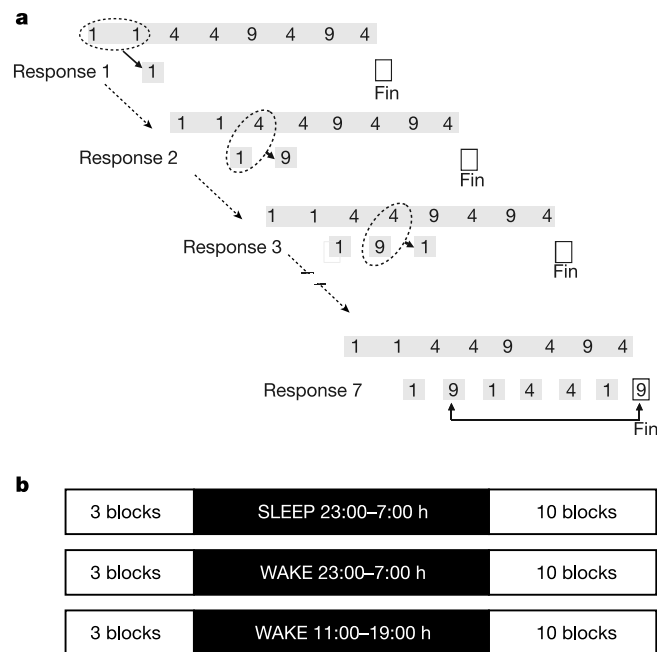


Figure 1 Task and experimental design. **a**, Number Reduction Task (NRT), illustrated by an example trial. On each trial, a different string of eight digits was presented. Each string was composed of the digits '1', '4', and '9'. For each string, subjects had to determine a digit defined as the 'final solution' of the task trial (Fin). This could be achieved by sequentially processing the digits pairwise from left to right according to two simple rules. One, the 'same rule', states that the result of two identical digits is just this digit (for example, '1' and '1' results in '1', as in response 1 here). The other rule, the 'different rule', states that the result of two non-identical digits is the remaining third digit of this three-digit system (for example, '1' and '4' results in '9' as in response 2 here). After the first response, comparisons are made between the preceding result and the next digit. The seventh response indicates the final solution, to be confirmed by pressing a separate key. Instructions stated that only this final solution was to be determined and this could be done at any time. Not mentioned to the subjects, the strings were generated in such a way that the last three responses always mirrored the previous three responses. This implies that in each trial the second response coincided with the final solution (arrow). Subjects who gain insight into this hidden rule abruptly cut short sequential responding by pressing the solution key immediately after the second response. **b**, Experimental design (main experiment). An 8-h period of nocturnal sleep, nocturnal wakefulness, or daytime wakefulness separated an initial training phase (three blocks) from later retesting (ten blocks).

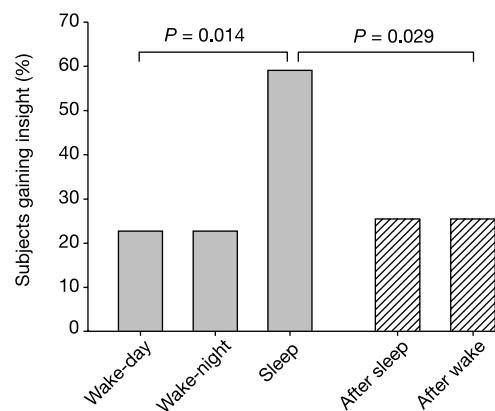


Figure 2 Effects of sleep and wakefulness on the occurrence of insight. Columns indicate percentage of subjects gaining insight into the hidden rule in the three experimental conditions of the main experiment (grey), in which subjects either slept (at night) or remained awake (at night or during daytime) between initial training and retesting, and in two supplementary conditions (hatched), where subjects were tested after nocturnal sleep or daytime wakefulness in the absence of initial training before these periods.

($F_{2,40} = 5.25$, $P < 0.01$, for solver/nonsolver $\times$ response type interaction). There were no similar differences between solvers and nonsolvers across the wake intervals, resulting generally in intermediate gains in reaction times ($P > 0.44$ and 0.14 , for respective main effect and interaction; Fig. 3b). Independent of the sleep/wake conditions, the first response in later solvers as the only response was delayed compared to nonsolvers already in the last block of the initial training ($F_{2,124} = 6.91$, $P < 0.005$, for response type $\times$ solver/nonsolver interaction). The slowing of the initial response might be related to processes of search and task analysis²¹, originating from an incipient representation of the hidden rule. Becoming particularly pronounced at retesting after sleep, this slowing could thus mean that sleep amplifies previously germinated antecedents of insight.

In conclusion, our results show that sleep acts on newly acquired mental representations of a task such that insight into hidden task structures is facilitated. Control conditions excluded the idea that this effect was due to non-specific effects of sleep deprivation, circadian rhythm, or proactive influences of sleep on subsequent capabilities of problem solving or divergent thinking^{22–24}. Our reaction time data argue against the view that the qualitative change evolves from a sleep-dependent strengthening of procedural (implicit) memories, which in sequential motor tasks expresses itself in distinctly accelerated reaction times²⁵. Here, sleep profoundly accelerated reaction times only in nonsolvers, but not in solvers, indicating that restructuring originates from an effect of sleep on memory representations different from those underlying procedural motor learning. Specifically, the slowing of reaction times in solvers appears to reflect the presence of an incipient representation of the rule overlapping with that required for implicit task performance. By amplification during sleep, this novel representation could eventually start to dominate the implicit memory representation, a process expressing itself in insightful behaviour as a consequence of an overall restructured representation.

The task representations associated with sleep-dependent gain of insight may be restructured by activity of the hippocampus and related medial temporal lobe structures, which, in connection with prefrontal cortical areas, are considered to play an essential role for generating awareness in memory^{26–28}. Reactivation of hippocampal cell assemblies during sleep^{6–8} is regarded as a mechanism by which recently encoded materials stored temporarily in autoassociative

hippocampal networks are played back to the neocortex where they are gradually incorporated into preexisting knowledge representations^{15,29}. This process of incorporation underlying long-term storage of previously acquired memories then forms the basis for a remodelling and qualitative restructuring of memory representations. Thus, our data support the concept that sleep, by hippocampal-neocortical replay, not only strengthens memory traces quantitatively, but can also 'catalyse' mental restructuring, thereby setting the stage for the emergence of insight. □

Methods

Sixty-six healthy subjects (age 18–31 yr) recruited at the University of Lübeck participated in the main experiment (twenty-two in each condition) and forty (age 20–32 yr) in the supplementary experiment (twenty in each condition). There were equal numbers of women and men in all five groups. The Number Reduction Task (explained in Fig. 1a) was adopted from ref. 18. Before the experiment proper, subjects had to perform without mistakes on ten practice strings to assure correct understanding of the 'same' and 'different' rule. Each task block consisted of 30 trials.

The hidden rule was abstract, that is, dependent on relational patterns rather than on fixed stimulus–stimulus or stimulus–response repetitions as in classical conditioning or in typical serial reaction-time tasks. In principle, insight into the hidden rule could be gained in different ways, which, however, were behaviourally equivalent and not treated separately here. The gain of insight reduced the total time to reach the final solution for a string abruptly from 8.73 ± 0.59 s to 2.39 ± 0.17 s. Post-experimental questionnaires confirmed the gain of explicit knowledge of the hidden rule in all subjects identified as solvers on the basis of their behavioural change.

In the design of the main experiment (Fig. 1b), three blocks were chosen as an initial training period based on previous work^{17,18} and on pilot experiments indicating that at this level of task difficulty only a few subjects were able to recognize the hidden rule within this early phase. Here, five subjects gained insight into the hidden rule already during initial training and, thus had to be replaced by additional subjects. To exclude that insight into the hidden rule had taken place spontaneously between initial training and retesting, subjects were asked in a control questionnaire before retesting whether any thoughts or mentations regarding the task had occurred after initial training. No subject reported any relevant thoughts or dreams pertinent to the task.

Subjects in the sleep condition of the main experiment spent a habituation night in the sleep laboratory before participation. Sleep was monitored by standard polysomnography including electroencephalogram (from left and right central electrodes), vertical and horizontal electrooculogram, and electromyogram (from chin electrodes). Recordings were scored off-line according to standard criteria³⁰, revealing normal sleep architecture for the experimental nights (total sleep time, 480.5 ± 3.9 min; sleep onset, 12.9 ± 1.5 min; wake, $0.2 \pm 0.1\%$; sleep stage 1, $4.5 \pm 0.7\%$; sleep stage 2, $58.6 \pm 1.8\%$; slow-wave sleep, $17.2 \pm 1.5\%$; rapid-eye-movement sleep, $17.7 \pm 1.2\%$).

Reaction time analyses across periods of sleep versus wakefulness (Fig. 3) were performed for correct response strings by a global 2 (sleep/wake) $\times$ 2 (solver/nonsolver) $\times$ 3 (response type) analysis of variance (ANOVA) and subsequent separate 2 (solver/nonsolver) $\times$ 3 (response type) ANOVA for sleep and wake subjects. Pairwise contrasts were specified by t -tests for statistically significant main effects and interactions. Degrees of freedom were adjusted using the Greenhouse–Geisser correction.

Received 21 August; accepted 17 November 2003; doi:10.1038/nature02223.

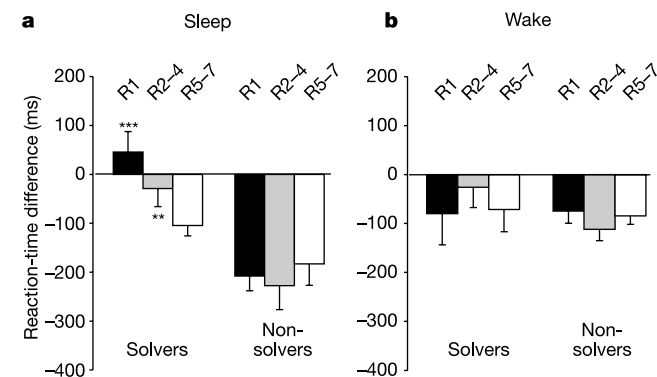


Figure 3 Reaction time analysis. **a**, **b**, Columns indicate changes in reaction times from the last block of initial training to the first block of retesting after eight hours of sleep (**a**) and wakefulness (**b**) for the initial response (R1, black), subsequent responses 2–4 undetermined by the task structure (R2–4, grey), and responses 5–7, which were determined by the task structure (R5–7, white), separately for solvers (who later gained insight) and nonsolvers (not gaining insight). *** $P < 0.001$, ** $P < 0.01$, for pairwise contrasts between solvers and nonsolvers.

- Köhler, W. *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology* (Liveright, New York, 1947).
- Sternberg, R. J. & Davidson, J. E. *The Nature of Insight* (Bradford Books/MIT Press, Cambridge, MA, 1995).
- Mazzarello, P. What dreams may come? *Nature* **408**, 523 (2000).
- Maquet, P. The role of sleep in learning and memory. *Science* **294**, 1048–1052 (2001).
- Smith, C. Sleep states and memory processes. *Behav. Brain Res.* **69**, 137–145 (1995).
- Buzsáki, G. Memory consolidation during sleep: a neurophysiological perspective. *J. Sleep Res.* **7** (Suppl. 1), 17–23 (1998).
- Pavides, C. & Winson, J. Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *J. Neurosci.* **9**, 2907–2918 (1989).
- Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994).
- Maquet, P. *et al.* Experience-dependent changes in cerebral activation during human REM sleep. *Nature Neurosci.* **3**, 831–836 (2000).
- Stickgold, R., Hobson, J. A., Fosse, R. & Fosse, M. Sleep, learning, and dreams: off-line memory reprocessing. *Science* **294**, 1052–1057 (2001).
- Fishbein, W. Disruptive effects of rapid eye movement sleep deprivation on long-term memory. *Physiol. Behav.* **6**, 279–282 (1971).
- Gais, S., Plihal, W., Wagner, U. & Born, J. Early sleep triggers memory for early visual discrimination skills. *Nature Neurosci.* **3**, 1335–1339 (2000).
- Plihal, W. & Born, J. Effects of early and late nocturnal sleep on declarative and procedural memory. *J. Cogn. Neurosci.* **9**, 534–547 (1997).
- Wagner, U., Gais, S. & Born, J. Emotional memory formation is enhanced across sleep intervals with high amounts of rapid eye movement sleep. *Learn. Mem.* **8**, 112–119 (2001).
- McNaughton, B. L., *et al.* in *Sleep and Brain Plasticity* (eds Maquet, P., Stickgold, R. & Smith, C.) 225–246 (Oxford Press, Oxford, UK, 2003).
- Thurstone, L. L. & Thurstone, T. G. Factorial studies of intelligence. *Psychomet. Monogr.* **2**, 94 (1941).
- Frensch, P., *et al.* in *Attention and Implicit Learning* (ed. Jiménez, L.) 335–366 (John Benjamins, Amsterdam/Philadelphia, 2002).

18. Rose, M., Haider, H., Weiller, C. & Büchel, C. The role of medial temporal lobe structures in implicit learning: an event-related fMRI study. *Neuron* **36**, 1221–1231 (2002).
19. Epstein, R., Kirshnit, C. E., Lanza, R. P. & Rubin, L. C. 'Insight' in the pigeon: antecedents and determinants of an intelligent performance. *Nature* **308**, 61–62 (1984).
20. Durso, F. T., Rea, C. B. & Dayton, T. Graph-theoretic confirmation of restructuring during insight. *Psychol. Sci.* **5**, 94–98 (1994).
21. Jiménez, L., Méndez, C. & Cleeremans, A. Comparing direct and indirect measures of sequence learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **22**, 948–969 (1996).
22. Cartwright, R. D. Problem solving: waking and dreaming. *J. Abnorm. Psychol.* **83**, 451–455 (1974).
23. Horne, J. A. Sleep loss and "divergent" thinking ability. *Sleep* **11**, 528–536 (1988).
24. Walker, M. P., Liston, C., Hobson, J. A. & Stickgold, R. Cognitive flexibility across the sleep-wake cycle: REM-sleep enhancement of anagram problem solving. *Brain Res. Cogn. Brain Res.* **14**, 317–324 (2002).
25. Fischer, S., Hallschmid, M., Elsner, A. L. & Born, J. Sleep forms memory for finger skills. *Proc. Natl Acad. Sci. USA* **99**, 11987–11991 (2002).
26. Clark, R. E. & Squire, L. R. Classical conditioning and brain systems: the role of awareness. *Science* **280**, 77–81 (1998).
27. Luo, J. & Niki, K. Function of hippocampus in "insight" of problem solving. *Hippocampus* **13**, 316–323 (2003).
28. McIntosh, A. R., Rajah, M. N. & Lobaugh, N. J. Functional connectivity of the medial temporal lobe relates to learning and awareness. *J. Neurosci.* **23**, 6520–6528 (2003).
29. McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
30. Rechtschaffen, A. & Kales, A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects* (NIH Publ. No. 204, Maryland, 1968).

Acknowledgements We thank S. Sabban, C. Benedict, M. Degirmenci, L. Hecking, A. Otterbein, and M. Rose for their assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft to J.B. and R.V.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to U.W. (wagner@kfg.uni-luebeck.de) or J.B. (born@kfg.uni-luebeck.de).

Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1

Mehrdad Matloubian^{1,2}, Charles G. Lo¹, Guy Cinamon¹, Matthew J. Lesneski¹, Ying Xu¹, Volker Brinkmann³, Maria L. Allende⁴, Richard L. Proia⁴ & Jason G. Cyster¹

¹Howard Hughes Medical Institute and Departments of Microbiology, Immunology and ²Medicine, University of California San Francisco, 513 Parnassus Avenue, San Francisco, California 94143-0414, USA

³Transplantation & Immunology, Novartis Institutes for BioMedical Research, WSJ-386.101, CH-4002 Basel, Switzerland

⁴Genetics of Development and Disease Branch, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, Maryland 20892-1821, USA

Adaptive immunity depends on T-cell exit from the thymus and T and B cells travelling between secondary lymphoid organs to survey for antigens. After activation in lymphoid organs, T cells must again return to circulation to reach sites of infection; however, the mechanisms regulating lymphoid organ exit are unknown. An immunosuppressant drug, FTY720, inhibits lymphocyte emigration from lymphoid organs, and phosphorylated FTY720 binds and activates four of the five known sphingosine-1-phosphate (S1P) receptors^{1–4}. However, the role of S1P receptors in normal immune cell trafficking is unclear. Here we show that in mice whose haematopoietic cells lack a single S1P receptor (S1P₁; also known as Edg1) there are no T cells in the periphery because mature T cells are unable to exit the thymus. Although B cells are present in peripheral lymphoid organs, they are severely deficient in blood and lymph. Adoptive cell transfer experiments

establish an intrinsic requirement for S1P₁ in T and B cells for lymphoid organ egress. Furthermore, S1P₁-dependent chemotactic responsiveness is strongly upregulated in T-cell development before exit from the thymus, whereas S1P₁ is downregulated during peripheral lymphocyte activation, and this is associated with retention in lymphoid organs. We find that FTY720 treatment downregulates S1P₁, creating a temporary pharmacological S1P₁-null state in lymphocytes, providing an explanation for the mechanism of FTY720-induced lymphocyte sequestration. These findings establish that S1P₁ is essential for lymphocyte recirculation and that it regulates egress from both thymus and peripheral lymphoid organs.

The G-protein-coupled receptors that are engaged by the lysophospholipid S1P are characterized most prominently for their functions in endothelial cells and for their roles in heart and vascular development⁵. However, S1P₁ and S1P₄ are also highly expressed in T and B lymphocytes, and sphingosine kinase is present in lymphoid organs^{6,7}. To test whether S1P₁ has an intrinsic role within lymphocytes, we generated fetal liver chimaeras, transplanting lethally irradiated wild-type mice with day 12.5 fetal liver cells from S1P₁ knockout donors⁸. Analysis of peripheral blood from reconstituted animals revealed an almost complete absence of peripheral T cells as well as reduced numbers of B cells in S1P₁-deficient (S1P₁^{−/−}) fetal liver chimaeras when compared with S1P₁^{+/+} control chimaeras (Fig. 1a, b, g). CD4 and CD8 T cells were also absent from spleen, lymph nodes and Peyer's patches (Fig. 1c), and were not found in liver or lungs (data not shown). In contrast with this peripheral T-cell deficiency, the thymus contained normal numbers of immature CD4 and CD8 double-positive thymocytes (Fig. 1d, f) but had an increased proportion of CD4 and CD8 single-positive cells (Fig. 1d). Further analysis of the single-positive populations revealed a marked increase in the number of mature L-selectin^{hi} cells but unchanged numbers of the immature L-selectin^{lo} cells (Fig. 1e, f; see also Supplementary Fig. 1). This thymic accumulation of mature single-positive T cells is reminiscent of that seen in mice expressing the G_i-inhibiting subunit of pertussis toxin within thymocytes⁹ and in animals treated with the immunosuppressive drug FTY720 (refs 2, 10).

In addition to high expression of L-selectin there were also greater numbers of cells expressing β7 integrin and Qa2 and expressing reduced levels of CD24 (Fig. 1e; see also Supplementary Fig. 1), additional phenotypic changes typical of mature single-positive thymocytes^{11,12}. CD69 was expressed at intermediate levels on the S1P₁^{−/−} single-positive thymocytes rather than being fully downregulated (Fig. 1e). In this regard it is notable that FTY720 treatment was recently found to downregulate CD69 on single-positive thymocytes¹⁰ and S1P₁ was identified in a screen for negative regulators of CD69 induction in activated Jurkat T cells¹³. Therefore, S1P₁ signalling may normally promote downregulation of CD69 in maturing single-positive thymocytes. By immunohistochemical analysis, thymic architecture appeared grossly normal in S1P₁^{−/−} fetal liver chimaeras (data not shown).

In contrast with the peripheral T-cell deficiency, S1P₁^{−/−} B cells were found in spleen, lymph nodes and Peyer's patches, and although they were altered in their proportions in these organs compared with wild-type controls, the total peripheral B-cell numbers were similar (Fig. 1h). The phenotype of peripheral B cells was normal with the exception that CD69 expression was elevated (Supplementary Fig. 2). In the bone marrow, pro/pre- and immature B cells were present at their usual frequency, whereas mature B cells that recirculate from blood to bone marrow were reduced (Fig. 1i). Therefore, peripheral B cells appeared to have accumulated in secondary lymphoid organs but seemed to be recirculating poorly. This possibility was tested further by quantification of lymphocyte numbers in lymph fluid isolated from the cisterna chyli (Fig. 1g and Methods). B-cell numbers were significantly reduced in lymph from the S1P₁^{−/−} fetal liver chimaeras

compared with control chimaeras (Fig. 1g). $S1P_1^{-/-}$ T cells were also absent from the lymph of $S1P_1^{-/-}$ chimaeras, as expected from their absence in the peripheral lymphoid organs, whereas they were plentiful in lymph from control chimaeras (data not shown).

To test whether T cells required $S1P_1$ for recirculation in the periphery, thymocytes from $S1P_1^{-/-}$ or wild-type fetal liver chimaeras were co-transferred with control CD45 congenic thymocytes into wild-type recipients. In previous studies it has been established that intravenously transferred single-positive thymocytes recirculate through peripheral lymphoid organs in a manner typical of mature T cells^{9,14}. One day after transfer, the number of cells in blood, lymphoid tissues and lymph of the recipient mice were enumerated. The transferred single-positive $S1P_1^{-/-}$ and wild-type T cells were readily able to enter secondary lymphoid organs (Fig. 2a, b), and they homed to T-cell compartments (Fig. 2c). However, $S1P_1^{-/-}$ cells disappeared from the blood and failed to appear in the lymph (Fig. 2a, b). Similar experiments were performed to examine B-cell recirculation, transferring CD45.2⁺ B cells that express the α allele of IgM and IgD (IgH^a) from $S1P_1^{-/-}$ or wild-type fetal liver chimaeras to CD45.1⁺ IgH^b wild-type recipients. Transferred $S1P_1^{-/-}$ B cells were found in all of the secondary

lymphoid organs, although in slightly altered proportions compared with the control cells, but were greatly reduced in numbers in blood and lymph (Fig. 2d). Immunohistochemical analysis of recipient spleens revealed that transferred $S1P_1^{-/-}$ B cells were present in lymphoid follicles but were mostly absent from the red pulp (Fig. 2c). These findings indicate that $S1P_1$ functions in egress of both T and B cells from peripheral lymphoid organs, including the splenic white pulp, with B cells showing slightly less dependence on this receptor than T cells.

The requirement for $S1P_1$ within thymocytes for their egress raised the possibility that exit from the thymus is controlled at the level of thymocyte $S1P_1$ expression and acquisition of $S1P$ responsiveness. As an approach to measure $S1P$ responsiveness of thymocyte subsets, wild-type thymocytes were tested for their ability to show a chemotactic response to $S1P$ (Fig. 3a, b). A robust chemotactic response was observed in the most mature subsets of CD4 and CD8 single-positive thymocytes, whereas little or no chemotactic response was detected in the more immature single-positive thymocytes, although background motility in the absence of attractant was sometimes higher (Fig. 3a, b). No chemotactic response to $S1P$ was detected in the CD4 and CD8 double-positive and the double-

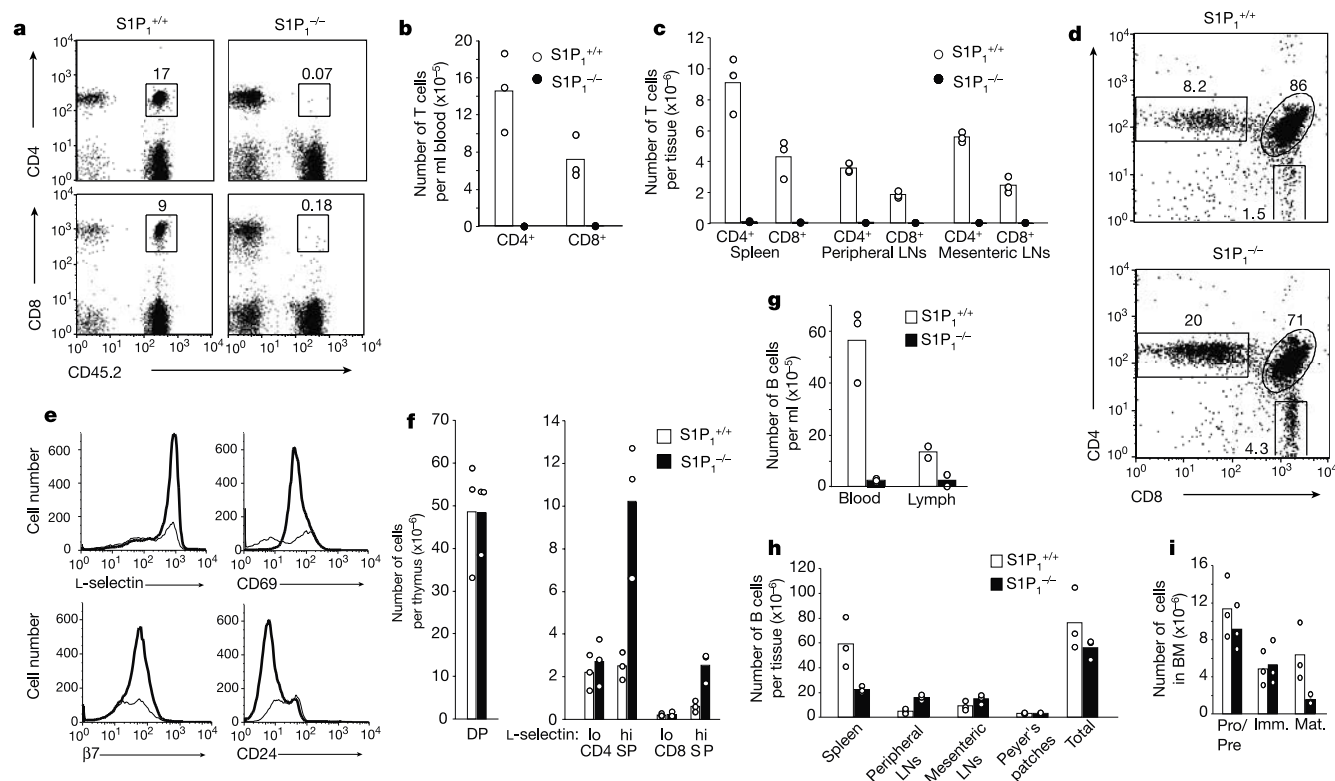


Figure 1 Defective emigration of $S1P_1^{-/-}$ mature thymocytes and recirculation of $S1P_1^{-/-}$ B cells. Irradiated CD45.1 mice were reconstituted with wild-type ($S1P_1^{+/+}$) or knockout ($S1P_1^{-/-}$) CD45.2 fetal livers and analysed 6–10 weeks later. **a**, Flow cytometric analysis of blood from $S1P_1^{+/+}$ and $S1P_1^{-/-}$ chimaeras stained for CD4 or CD8 and CD45.2. Numbers show the per cent total lymphocyte-sized cells. **b, c**, Number of circulating $S1P_1^{+/+}$ and $S1P_1^{-/-}$ T cells in blood (**b**) and peripheral lymphoid tissues (**c**). Bars indicate averages and circles indicate values for individual mice ($n = 3$). **d**, Flow cytometric analysis of thymocytes from $S1P_1^{+/+}$ and $S1P_1^{-/-}$ chimaeras stained to detect CD4, CD8 and CD45.2. Profiles shown are gated on CD45.2⁺ donor-derived cells, which comprised >98% of total thymocytes. Numbers shown are per cent of CD45.2⁺ cells. **e**, Expression of maturation markers L-selectin (CD62L), CD69, integrin $\beta 7$ and CD24 (heat-stable antigen) on $S1P_1^{+/+}$ (thin lines) and $S1P_1^{-/-}$ (thick lines) CD4 single-positive thymocytes. Similar numbers of total thymocytes were analysed for the wild-type and

knockout samples and the higher value of the histograms for $S1P_1^{-/-}$ samples reflects the greater number of single-positive cells. **f**, Number of CD4, CD8 double-positive (DP) and L-selectin^{lo} (immature) or L-selectin^{hi} (mature) single-positive (SP) thymocytes in $S1P_1^{+/+}$ or $S1P_1^{-/-}$ chimaeras ($n = 3$). **g, h**, Numbers of donor-derived CD45.2⁺ $S1P_1^{+/+}$ or $S1P_1^{-/-}$ CD19⁺ B cells in blood and lymph (**g**) or secondary lymphoid organs (**h**) of fetal liver chimaeras. Total in **h** represents the sum of B-cell numbers shown in the indicated tissues ($n = 3$). **i**, Bone marrow cells from two femurs and tibias of each fetal liver chimaera were enumerated and stained for CD45.2, B220, IgM and IgD. Bars show number of CD45.2⁺ B220⁺ pro/pre-B cells (Pro/pre; IgM⁻, IgD⁻), immature B cells (Imm.; IgM⁺, IgD⁻) and mature recirculating B cells (Mat.; IgM⁺, IgD⁺) ($n = 3$). More than 98% of total B220⁺ bone marrow cells were CD45.2⁺ donor-derived cells.

negative populations (Fig. 3b). Peripheral T cells also exhibited a chemotactic response to S1P (not shown), consistent with previous findings⁷. The thymocyte response was predominantly chemotactic rather than chemokinetic as cells migrated only poorly when S1P was added without a gradient. By quantitative polymerase chain reaction (PCR) analysis, a strong (approximately 50-fold) upregulation in S1P₁ expression occurred between the double-positive and CD4 and CD8 single-positive stages of thymocyte maturation (Fig. 3c and data not shown), and analysis of the immature and mature CD4 single-positive cells revealed an additional 30-fold upregulation on the most mature single-positive cells such that the cells had levels equivalent to peripheral T cells (Fig. 3c and data not shown). Thymocytes also expressed S1P₂ and S1P₄, although expression was not upregulated between the double-positive and single-positive stages (Fig. 3c). The chemokine receptor CCR7 was increased during the double-positive to single-positive transition, as expected, but was not further upregulated during the immature to mature single-positive transition (Fig. 3c). Notably, S1P₁^{-/-} thymocytes failed to display a chemotactic response to S1P *in vitro*, while continuing to migrate to the CCR7 ligand (CCL21; Fig. 3d, e), establishing that S1P₁ is the major S1P receptor on these cells that supports chemotaxis.

During the immune response, antigen-specific lymphocytes are transiently retained within antigen-bearing lymphoid organs, undergoing activation and clonal expansion and then exiting as effector cells^{15,16}. To examine whether this retention mechanism might involve transient downregulation of S1P responsiveness and of the S1P₁ receptor on the activated T cells, as has been observed for *in vitro* activated cells⁷, ovalbumin-specific DO11.10 T-cell antigen

receptor (TCR) transgenic T cells were transferred to wild-type hosts that were then immunized with ovalbumin. At 0, 1 and 3 days after antigen exposure, draining lymph node cells were isolated and used in S1P chemotaxis assays or for RNA isolation and quantitative PCR analysis of S1P₁ levels. One day after antigen exposure, the activated antigen-specific T cells had lost their responsiveness to S1P and they had downregulated S1P₁ expression 100-fold (Fig. 4a, b). Three days after immunization, many recently divided antigen-specific T cells had appeared in circulation (data not shown), and at this time the activated draining lymph node cells exhibited restored S1P responsiveness and increased S1P₁ receptor levels (Fig. 4a, b). Therefore, downregulation of S1P₁ responsiveness is associated with the initial retention of activated T cells in lymphoid organs, and reacquisition of responsiveness is associated with their exit.

The similar effects of S1P₁ deficiency and treatment with FTY720 on thymocyte emigration and lymphocyte recirculation suggested that active doses of FTY720 might cause functional inactivation of S1P₁ on lymphocytes. To test this possibility, mice were treated with FTY720 and 14 h later—a time point when lymphocyte numbers are markedly reduced in blood (Supplementary Fig. 3a)^{1,3} and lymph (Supplementary Fig. 3b), and when thymocyte emigration is already affected (Supplementary Fig. 3c)^{2,10}—cells were isolated and tested for S1P responsiveness. Mature single-positive thymocytes and peripheral T cells from FTY720-treated mice failed to migrate in response to S1P (Fig. 5a, b) while migrating normally to CCL21/SLC (data not shown), demonstrating that *in vivo* exposure to this drug, or its phosphorylated form, inactivates the ability of S1P₁ to support T-lymphocyte chemotaxis. Flow cytometric analysis of cells expressing an epitope-tagged S1P₁ receptor, before and after incu-

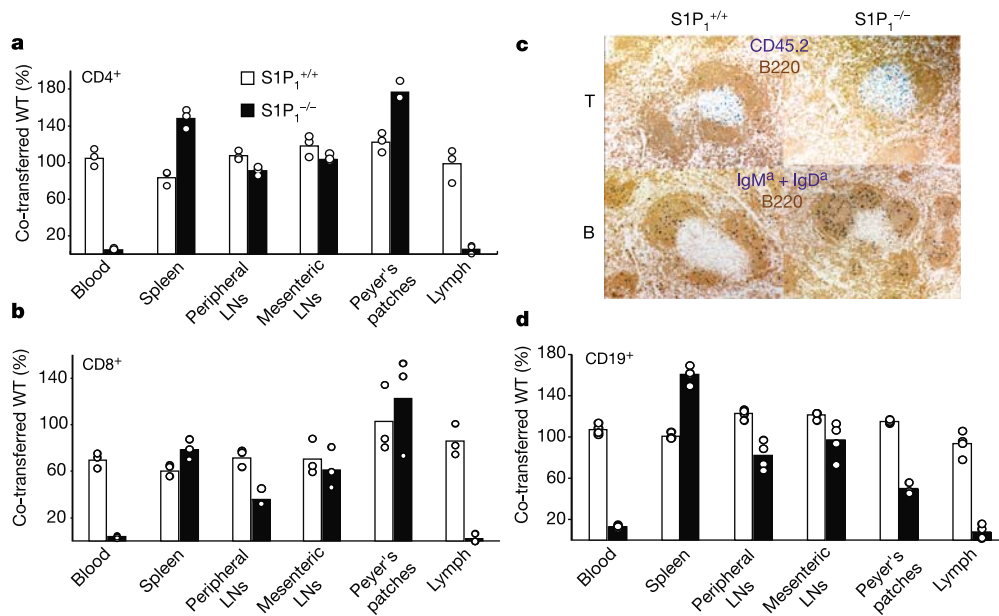


Figure 2 Transferred S1P₁^{-/-} T and B cells accumulate in secondary lymphoid organs and fail to exit. **a**, **b**, CD45.2⁺ thymocytes from S1P₁^{+/+} or S1P₁^{-/-} fetal liver chimaeras were co-transferred with approximately equal numbers of CMTMR-labelled CD45.1⁺ wild-type thymocytes into CD45.1 B6 mice. After 24 h, blood, lymph and peripheral lymphoid tissues were examined for the presence of transferred cells by flow cytometry. Frequency of transferred S1P₁^{+/+} or S1P₁^{-/-} CD4 single-positive (**a**) and CD8 single-positive (**b**) T cells in the blood, secondary lymphoid organs and lymph of recipients 24 h after transfer, as a per cent of the CMTMR-labelled co-transferred control cells (see Methods). LNs, lymph nodes. **c**, Top panel: spleen sections from CD45.1 B6 mice 24 h after receiving CD45.2⁺ S1P₁^{+/+} or S1P₁^{-/-} thymocytes as described in **a**, stained to detect transferred CD45.2⁺ T cells (blue) and endogenous B220⁺ B cells (brown). Bottom

panel: spleen sections from Igh^b mice 24 h after receiving Igh^{a+} spleen and lymph node cells from S1P₁^{+/+} or S1P₁^{-/-} mice, stained to detect transferred Igh^{a+} B cells (blue) and endogenous B220⁺ B cells (brown). Objective magnification $\times 5$. **d**, CD45.2⁺ S1P₁^{+/+} or S1P₁^{-/-} B cells were co-transferred with equal numbers of CMTMR-labelled CD45.1⁺ wild-type B cells into CD45.1 B6 mice. Frequency of transferred CD19⁺ B cells in blood, secondary lymphoid tissues and lymph of the recipients 40 h after the transfer was determined as for T cells in **a**, **b**. For transfer experiments, the average and individual values for 3–4 recipients per group is shown. The data are representative of three experiments using chimaeras made from at least two different S1P₁^{+/+} and S1P₁^{-/-} fetal livers.

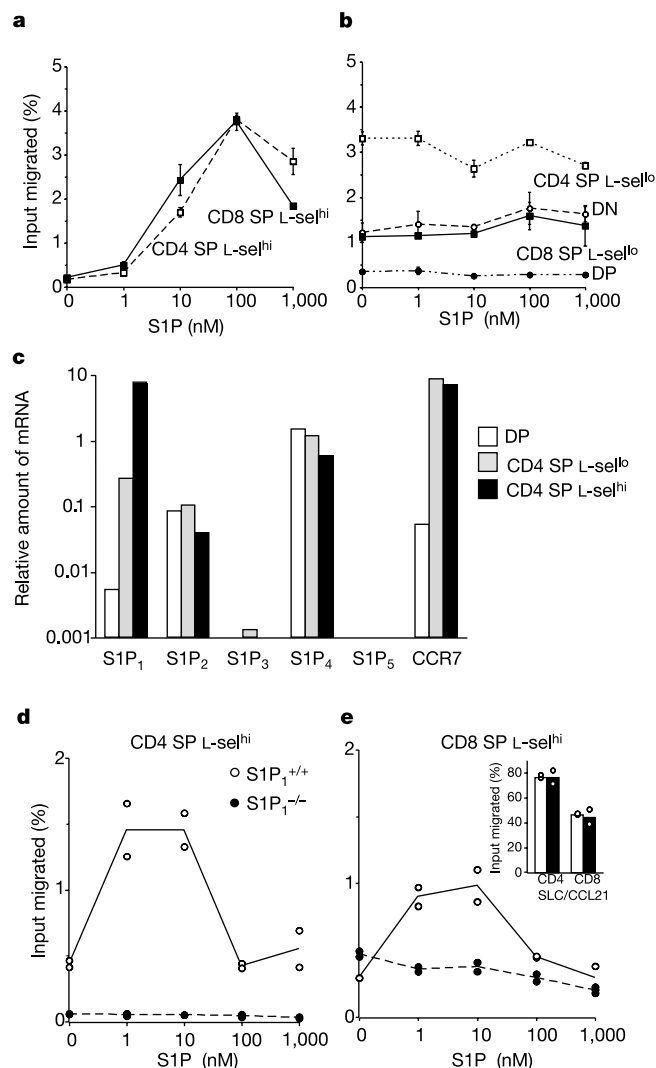


Figure 3 Mature single-positive (SP) thymocytes upregulate S1P₁ and respond to S1P in an S1P₁-dependent manner. **a, b**, Chemotactic response of different thymocyte populations to S1P. The percentage of input cells of the indicated population that migrated to S1P is shown with the error bars representing the range of response for duplicate samples. **c**, Quantitative PCR analysis of S1P receptors and CCR7 in sorted thymocyte subsets, expressed as relative amount of indicated messenger RNA normalized to HPRT. Representative of three experiments. DP, double positive. **d, e**, Chemotaxis of S1P₁^{-/-} and S1P₁^{+/+} mature (L-selectin^{hi}) CD4 single-positive and CD8 single-positive thymocytes to S1P. The percentage of the input cell population that responded to S1P is shown in duplicate and the lines are drawn through the mean values. Inset shows chemotactic response of S1P₁^{+/+} and S1P₁^{-/-} CD4 single-positive and CD8 single-positive thymocytes to 1 μ g ml⁻¹ CCL21/SLC.

bation with FTY720, revealed that the drug caused almost complete downmodulation of the receptor (Fig. 5c). The phosphorylated form of FTY720 was also active in downregulating S1P₁ (Supplementary Fig. 4), and we speculate that the non-phosphorylated FTY720 became phosphorylated by sphingosine kinase in the cells before inducing receptor downmodulation. FTY720-induced inactivation of lymphocyte S1P₁ can therefore explain the egress-blocking effects of this drug.

These observations establish that S1P₁ is required within maturing thymocytes for emigration from the thymus and within lymphocytes for their egress from spleen, lymph nodes and Peyer's patches. S1P₁ couples to G_i¹⁷ and it seems likely that the essential role of G_i in T-cell egress from the thymus⁹ is at least in part due to a

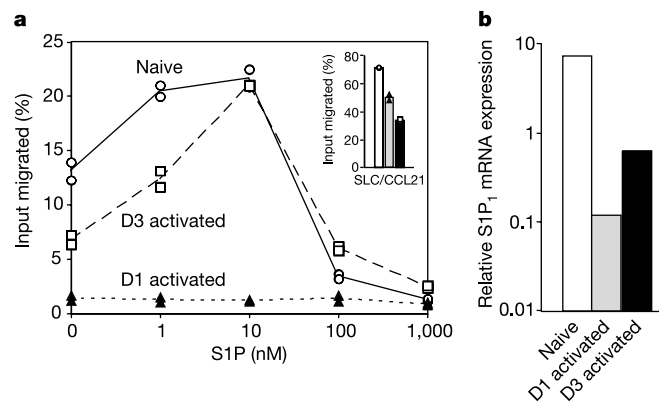


Figure 4 *In vivo* T-cell activation is associated with loss and subsequent reacquisition of S1P₁ and S1P responsiveness. **a**, Chemotactic response of naive CD4 T cells (circles) or TCR transgenic DO11.10 CD4 T cells one day (D1, triangles) or three days (D3, squares) after *in vivo* activation with antigen. The percentage of the input cell population that responded to the indicated concentration of S1P is shown in duplicate. Inset shows chemotactic response of the same populations to 1 μ g ml⁻¹ CCL21/SLC. **b**, Quantitative PCR analysis of S1P₁ mRNA expression, normalized to HPRT mRNA levels, in sorted naive CD4 T cells (white bar), and DO11.10 CD4 T cells 1 day (grey bar) or 3 days (black bar) after *in vivo* activation with antigen. Representative of three experiments. For details of markers used to identify and sort activated T cells, see Methods.

requirement in S1P₁ signalling. Regulated expression of S1P₁ within thymocytes seems to be a mechanism controlling the timing of their exit from the thymus. We speculate that in the periphery, the functional activity of S1P₁ in T and B lymphocytes is regulated in a manner that influences the duration of their transit through secondary lymphoid organs. Precisely how S1P₁ functions to facilitate egress needs further investigation, but it is worth noting that S1P levels are constitutively high within blood (100–400 nM in plasma^{18–20}). S1P may therefore function to attract cells out of these organs. An alternative mechanism might involve S1P signalling transiently inactivating responsiveness to attractive lymphoid chemokines²¹ or other yet unidentified retention signals. Our findings also indicate that downregulation of S1P₁ expression in activated lymphocytes will contribute to promoting their retention within the antigen-bearing lymphoid organ during the initial phase of the adaptive immune response. Similarly, the downregulation of S1P responsiveness reported to occur in dendritic cells during maturation²² might contribute to the efficient trapping of these cells within lymph nodes. It will be important to examine S1P₁ levels on lymphoma cells to determine whether they relate to retention or release of the cells from lymphoid organs, and to define whether changes in S1P production during immune responses are associated with lymphoid organ shutdown¹⁵. Finally, our findings indicate that of the four receptors engaged by the phosphorylated form of FTY720, its effect on lymphocyte egress can be explained by a selective action on S1P₁, and that this effect is on the lymphocyte rather than the endothelium or other S1P₁-expressing cell types. Although FTY720 phosphate has been demonstrated to have S1P₁ agonistic activity^{3,4}, our results indicate that its egress-blocking effect is through S1P₁ inactivation. The differing effects of FTY720 treatment and S1P₁ deficiency on CD69 expression suggest that FTY720 phosphate may act *in vivo* as a partial S1P₁ agonist or cause an acute positive signal before inactivating the receptor. In light of these findings, we propose that a selective S1P₁ antagonist may be a highly effective immunosuppressant. □

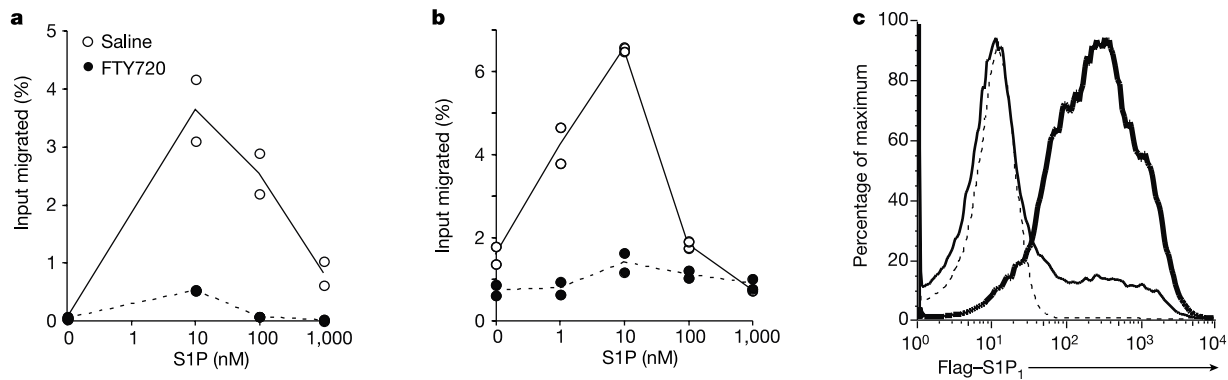


Figure 5 FTY720 treatment uncouples S1P₁ from the chemotactic machinery and downmodulates S1P₁ surface expression. **a**, **b**, *Ex vivo* chemotaxis assays of L-selectin^{hi} CD4 single-positive thymocytes (**a**) or peripheral lymph node CD4 T cells (**b**) from mice 14 h after treatment with FTY720 (filled circles) or saline (open circles). The percentage of the input cell population that responded to S1P is shown in duplicate and the lines are

drawn through the mean values. **c**, A murine B-cell line (WEHI231) expressing Flag-tagged S1P₁ was incubated with either saline (thick line) or FTY720 (thin line) for 8 h and then stained with an antibody to Flag to determine surface Flag-S1P₁ expression. Vector (dashed line) or Flag-S1P₁ transduced cells were identified by expression of an independent marker (see Methods).

Methods

Mice, fetal liver chimaeras and adoptive transfers

B6 and B6 CD45.1 mice were from National Cancer Institutes. S1P₁^{+/−} mice were generated as described⁸ and were crossed for at least two generations to C57BL/6 (B6) mice and then S1P₁^{+/−} mice were intercrossed. Three different S1P₁^{+/−} males and five separate S1P₁^{+/−} females were used to generate the S1P₁ wild-type and S1P₁ knockout fetuses used in this study. As S1P₁^{−/−} fetuses die by E14.5 (ref. 8), fetal liver cells were prepared at E12.5, genotyped by PCR⁸ and each preparation was used to reconstitute about three lethally irradiated B6 CD45.1 mice²³. Mice were analysed 6–10 weeks after reconstitution.

For T-cell adoptive transfers, B6 CD45.1 thymocytes were labelled with 1 μM 5-(and-6)-((4-chloromethyl)benzoyl)amino)tetramethylrhodamine (CMTMR, Molecular Probes) at 37 °C for 30 min, washed, and mixed with total thymocytes from either S1P₁^{+/−} or S1P₁^{−/−} fetal liver chimaeras containing comparable numbers of L-selectin^{hi} single-positive cells, and injected intravenously into B6 CD45.1 recipients. At 24 h after transfer, more than 95% of the transferred cells identified in lymphoid organs and blood were CD4 or CD8 single-positive cells. For B-cell transfers, B6 CD45.1 spleen and lymph node cells were labelled with CMTMR and mixed with spleen and lymph node cells from S1P₁^{+/−} or S1P₁^{−/−} fetal liver chimaeras in numbers normalized for B-cell content, and injected intravenously into B6 CD45.1 mice. After 24–40 h, blood, lymph and lymphoid tissues from recipients were analysed by flow cytometry. The frequency of the CD45.2⁺ or CMTMR⁺ transferred cells of each lineage in lymphoid tissues was determined by staining for CD45.2 and lineage markers (CD4, CD8, and CD19). For each lineage, the relative frequency of knockout or wild-type cells to co-transferred control cells in any given tissue was determined by dividing the per cent of that lineage which were transferred knockout or wild-type cells by the per cent of cells that were transferred CMTMR⁺. The frequency was corrected for differences in the input ratio of knockout or wild-type cells and CMTMR⁺ control cells. For *in vivo* activation of T cells, 1–2 × 10⁷ spleen and lymph node cells from BALB/c DO11.10 mice (Jackson Laboratories), corresponding to approximately 3 × 10⁶ TCR transgenic T cells, were transferred to BALB/c (Jackson Laboratories) recipients. In some cases the DO11.10 donor cells were labelled with 5 μM 5- and 6-carboxyfluorescein diacetate succinimidyl ester (CFSE, Molecular Probes) for 20 min at 37 °C. The day after cell transfer (day 0) mice were immunized subcutaneously with a total of 200 μg ovalbumin (Sigma) emulsified in complete Freund's adjuvant (Sigma) at several sites along the back. For FTY720 treatment, B6 mice were injected intraperitoneally with 0.1 mg kg^{−1} or 1.0 mg kg^{−1} of the drug for thymus or peripheral lymphoid tissue analysis, respectively, or an equivalent volume of saline.

Chemotaxis assays

For chemotaxis assays, cells transmigrated across 5-μm transwell filters (Corning Costar Corp.) for 3 h to S1P (Sigma), chemokines (R&D Systems) or medium in the bottom chamber, and were enumerated by flow cytometry²³. Transwell assays were performed in duplicate for each S1P or chemokine concentration, and were repeated using cells from a minimum of three different animals of each type.

Flow cytometry and immunohistochemistry

Flow cytometric analysis was on a FACSCaliber (Becton Dickinson). All antibodies were from Pharmingen with the exception of biotinylated Flag M2 (Sigma). Immunohistochemical analysis was as described²³.

Cell sorting and quantitative PCR

Thymocytes from 4–6-week-old B6 mice were stained with CD4-APC, CD8-PE, L-selectin-biotin and CD69 FITC followed by streptavidin PE-Cy7, dead cells were excluded by propidium iodide staining, and cells were isolated using a MoFlo sorter (Cytomation). Transferred DO11.10 TCR transgenic cells were sorted from pooled

axillary, brachial and inguinal lymph nodes of control and immunized mice. KJ1.26⁺ CD4⁺ TCR transgenic cells were identified as activated by low expression of L-selectin (day 1) or dilution of CFSE label to an extent indicative of greater than three cell divisions (day 3). Non-transgenic naive L-selectin^{hi} CD4⁺ T cells were sorted in parallel at day 1. RNA was prepared using RNeasy (Qiagen), and equivalent amounts of complementary DNA were used in quantitative PCR on an ABI 7700 sequence detection instrument (Applied Biosystems) with the following primers/probes (Integrated DNA Technologies): S1P₁ forward primer GTGTAGACCCAGAGTCCTGCG, reverse primer AGCTTTTCCTGGCTGGAGAG, probe CGGCTTGAGCGAGGCTGCTGT; S1P₂ forward primer GGCCTAGCCAGTGCTCAGC, reverse primer CCTTGGTGTAATTG TAGTGTGCCAGA, probe CAGAGTACCTCAATCCTGA; S1P₃ forward primer GGAGCCCTAGACGGGAGT, reverse primer CCGACTGCGGGAAGAGTGT, probe AGAACGAGAGCCTATTTT; S1P₄ forward primer CCTGGAACCTCACTTTATAGACC AGG, reverse primer AGAAAGCGTGCCATAGGAC, probe TGGCCTCGAAGTACG AAATCCGCC; S1P₅ forward primer GAGTGCCGGTTACAGGAGACTT, reverse primer CGCTGTGTGTCCTGCC, probe CACAGTGTCCAGTGGG; CCR7 forward primer CAGCCTTCCTGTGTGATTCTACA, reverse primer ACCACGACAGCTTTTCTCT, probe AGAGCACCATGGACCCAGGAAAC; HPRT forward primer AGGTTGCAAGCTTGCTGGT, reverse primer TGAAGTACTCATTATAGTCAAGGCA, probe TGTGTGATACAGGCCAGACTTTGTTGGAT.

Lymph collection

After the mice were killed, the peritoneal cavity was washed with 6 ml of RPMI medium and then its contents were exposed ventrally. About 500 μl of blood was aspirated from the inferior vena cava. Under a stereomicroscope, the cisterna chyli was identified as a small cream-coloured sac located dorsal to the left renal lymph node. A fine borosilicate glass microcapillary pipette (Sutter Instrument) was then inserted into the cisterna chyli and lymph fluid was removed and the volume measured with a micropipette. Lymph cells were stained and the entire sample was analysed by flow cytometry. Cell numbers determined in this manner were divided by the volume of collected lymph to determine the concentration.

Epitope-tagged S1P₁

The full-length murine S1P₁ open reading frame together with an in-frame pre-prolactin leader and NH₂-terminal Flag-epitope was cloned into the MSCV2.2 retroviral vector, and retrovirus-containing culture supernatant was generated using the Bosc23 packing cell line as described²⁴. This vector also contains an IRES-tagged human CD4, which can be used as an independent surface marker to identify transduced cells. WEHI231 B cells were spin-infected and 65 h later incubated with 1 μM FTY720, its phosphorylated form (FTY720-P), or equivalent volumes of saline or vehicle for 8 h, followed by flow cytometric analysis for surface Flag expression.

Received 3 November; accepted 12 December 2003; doi:10.1038/nature02284.

- Chiba, K. *et al.* FTY720, a novel immunosuppressant, induces sequestration of circulating mature lymphocytes by acceleration of lymphocyte homing in rats. I. FTY720 selectively decreases the number of circulating mature lymphocytes by acceleration of lymphocyte homing. *J. Immunol.* **160**, 5037–5044 (1998).
- Yagi, H. *et al.* Immunosuppressant FTY720 inhibits thymocyte emigration. *Eur. J. Immunol.* **30**, 1435–1444 (2000).
- Mandala, S. *et al.* Alteration of lymphocyte trafficking by sphingosine-1-phosphate receptor agonists. *Science* **296**, 346–349 (2002).
- Brinkmann, V. *et al.* The immune modulator FTY720 targets sphingosine 1-phosphate receptors. *J. Biol. Chem.* **277**, 21453–21457 (2002).
- Allende, M. L. & Proia, R. L. Sphingosine-1-phosphate receptors and the development of the vascular system. *Biochim. Biophys. Acta* **1582**, 222–227 (2002).
- Kohama, T. *et al.* Molecular cloning and functional characterization of murine sphingosine kinase.

- J. Biol. Chem.* **273**, 23722–23728 (1998).
- Graeber, M. & Goetzl, E. J. Activation-regulated expression and chemotactic function of sphingosine 1-phosphate receptors in mouse splenic T cells. *FASEB J.* **16**, 1874–1878 (2002).
 - Liu, Y. *et al.* Edg-1, the G protein-coupled receptor for sphingosine-1-phosphate, is essential for vascular maturation. *J. Clin. Invest.* **106**, 951–961 (2000).
 - Chaffin, K. E. & Perlmutter, R. M. A pertussis toxin sensitive process controls thymocyte emigration. *Eur. J. Immunol.* **21**, 2565–2573 (1991).
 - Rosen, H., Alfonso, C., Surh, C. D. & McHeyzer-Williams, M. G. Rapid induction of medullary thymocyte phenotypic maturation and egress inhibition by nanomolar sphingosine 1-phosphate receptor agonist. *Proc. Natl Acad. Sci. USA* **100**, 10907–10912 (2003).
 - Lucas, B., Vasseur, F. & Penit, C. Production, selection, and maturation of thymocytes with high surface density of TCR. *J. Immunol.* **153**, 53–62 (1994).
 - Gabor, M. J., Godfrey, D. I. & Scollay, R. Recent thymic emigrants are distinct from most medullary thymocytes. *Eur. J. Immunol.* **27**, 2010–2015 (1997).
 - Chu, P. *et al.* Systematic identification of regulatory proteins critical for T-cell activation. *J. Biol.* **2**, 211–2116 (2003).
 - Parrot, D. M. V. & de Sousa, M. Thymus-dependent and thymus-independent populations: origin, migratory patterns and lifespan. *Clin. Exp. Immunol.* **8**, 663–673 (1971).
 - Hall, J. G. & Morris, B. The immediate effect of antigens on the cell output of a lymph node. *Br. J. Exp. Pathol.* **46**, 450–454 (1965).
 - Sprent, J., Miller, J. F. A. P. & Mitchell, G. F. Antigen-induced selective recruitment of circulating lymphocytes. *Cell. Immunol.* **2**, 171–181 (1971).
 - Hla, T. Sphingosine 1-phosphate receptors. *Prostaglandins Lipid Mediat.* **64**, 135–142 (2001).
 - Edsall, L. C. & Spiegel, S. Enzymatic measurement of sphingosine 1-phosphate. *Anal. Biochem.* **272**, 80–86 (1999).
 - Murata, N. *et al.* Interaction of sphingosine 1-phosphate with plasma components, including lipoproteins, regulates the lipid receptor-mediated actions. *Biochem. J.* **352**, 809–815 (2000).
 - Caligan, T. B. *et al.* A high-performance liquid chromatographic method to measure sphingosine 1-phosphate and related compounds from sphingosine kinase assays and other biological samples. *Anal. Biochem.* **281**, 36–44 (2000).
 - Graeber, M., Shankar, G. & Goetzl, E. J. Cutting edge: suppression of T cell chemotaxis by sphingosine 1-phosphate. *J. Immunol.* **169**, 4084–4087 (2002).
 - Idzko, M. *et al.* Sphingosine 1-phosphate induces chemotaxis of immature and modulates cytokine-release in mature human dendritic cells for emergence of Th2 immune responses. *FASEB J.* **16**, 625–627 (2002).
 - Hargreaves, D. C. *et al.* A coordinated change in chemokine responsiveness guides plasma cell movements. *J. Exp. Med.* **194**, 45–56 (2001).
 - Reif, K. *et al.* Balanced responsiveness to chemoattractants from adjacent zones determines B-cell position. *Nature* **416**, 94–99 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to C. Low for technical assistance; C. McArthur and S. Jiang for cell sorting; R. Albert at Novartis Institutes for BioMedical Research for synthesis of FTY720-P; and T. Okada, C. Allen and S. Watson for helpful discussions. M.M. is supported by the Pfizer Postdoctoral Fellowship in Immunology and Rheumatology and the Rosalind Russell Medical Research Center for Arthritis at University of California, San Francisco; J.G.C. is a Packard fellow and an HHMI assistant investigator. This work was supported in part by grants from the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.G.C. (cyster@itsa.ucsf.edu).

The mitochondrial calcium uniporter is a highly selective ion channel

Yuriy Kirichok, Grigory Krapivinsky & David E. Clapham

Howard Hughes Medical Institute, Department of Cardiovascular Research, Children's Hospital, and Department of Neurobiology, Harvard Medical School, 320 Longwood Avenue, Boston, Massachusetts 02115, USA

During intracellular Ca^{2+} signalling mitochondria accumulate significant amounts of Ca^{2+} from the cytosol^{1,2}. Mitochondrial Ca^{2+} uptake controls the rate of energy production^{1,3,4}, shapes the amplitude and spatio-temporal patterns of intracellular Ca^{2+} signals^{1,5–8}, and is instrumental to cell death^{9,10}. This Ca^{2+} uptake is undertaken by the mitochondrial Ca^{2+} uniporter (MCU) located in the organelle's inner membrane^{11,12}. The uniporter passes Ca^{2+} down the electrochemical gradient maintained

across this membrane without direct coupling to ATP hydrolysis or transport of other ions¹¹. Carriers are characterized by turn-over numbers that are typically 1,000-fold lower than ion channels, and until now it has been unclear whether the MCU is a carrier or a channel¹³. By patch-clamping the inner mitochondrial membrane, we identified a previously unknown Ca^{2+} -selective ion channel sensitive to inhibitors of mitochondrial Ca^{2+} uptake. Our data indicate that this unique channel binds Ca^{2+} with extremely high affinity (dissociation constant ≤ 2 nM), enabling high Ca^{2+} selectivity despite relatively low cytoplasmic Ca^{2+} concentrations. The channel is inwardly rectifying, making it especially effective for Ca^{2+} uptake into energized mitochondria. Thus, we conclude that the properties of the current mediated by this novel channel are those of the MCU.

Since its discovery in 1961 (ref. 14), mitochondrial Ca^{2+} uptake has been measured in suspensions of isolated mitochondria. This approach does not reliably control ion concentrations or voltage gradients across the inner mitochondrial membrane, resulting in inaccuracies in experimental results^{11,12}. The patch-clamp method solves these problems but its implementation in mitochondria is a severe technical challenge. Here we patch-clamp single mitoplasts (2–5- μm vesicles of inner mitochondrial membrane) isolated from COS-7 cells (Fig. 1a, b), to measure directly whole-mitoplast and single-channel Ca^{2+} currents.

In the whole-mitoplast configuration, voltage ramps from -160 mV (potential across the inner membrane of energized mitochondria) to $+80$ mV elicited an inwardly rectifying current that gradually increased as free Ca^{2+} concentration at the cytosolic surface of the inner membrane ($[\text{Ca}^{2+}]_c$) was varied from $20 \mu\text{M}$ to $100 \mu\text{M}$. These concentrations are comparable to $[\text{Ca}^{2+}]$ in microdomains near endoplasmic/sarcoplasmic reticulum Ca^{2+} release or plasma membrane Ca^{2+} channels, sensed by neighbouring mitochondria^{1,2,15,16}. At $100 \mu\text{M}$ $[\text{Ca}^{2+}]_c$, the inwardly rectifying current reached amplitudes of 20–30 pA (Fig. 1c). As mitoplasts are small, this amplitude corresponds to a significant current density ($55 \pm 19 \text{ pA pF}^{-1}$ at -160 mV, $n = 4$). To examine the level at which the current saturates, $[\text{Ca}^{2+}]_c$ was elevated from $20 \mu\text{M}$ to 105 mM (Fig. 1c–e). Notably, the Ca^{2+} carrying capacity of the current was enormous, reaching half-saturation at approximately 20 mM $[\text{Ca}^{2+}]_c$. Elimination of Na^+ in the bath did not reduce the current, suggesting that Ca^{2+} was the primary charge carrier (not shown). We refer to the mitoplast inwardly rectifying Ca^{2+} current as I_{MiCa} (for mitochondrial Ca^{2+} current).

I_{MiCa} elicited by voltage steps rapidly inactivated to a plateau level (Fig. 1f). The time-independent plateau was the major component of the current with the transient component averaging $21 \pm 14\%$ of the net peak current ($n = 16$). I_{MiCa} amplitude was not noticeably altered when pipette (intramitoplast) $[\text{Ca}^{2+}]$ was varied from <10 nM to about $10 \mu\text{M}$ (no EGTA or EDTA). Thus, unlike many Ca^{2+} -selective currents, I_{MiCa} lacked Ca^{2+} -dependent inactivation at energized mitochondrial potentials (approximately -160 mV).

Similar to MCU-mediated mitochondrial Ca^{2+} uptake^{11,17–19}, Ruthenium 360 (Ru360) inhibited I_{MiCa} in a dose-dependent manner (half-maximal inhibitory concentration (IC_{50}), 2 nM) and was a more potent inhibitor than ruthenium red (RuR, IC_{50} 9 nM ; Fig. 2).

RuR-sensitive (200 nM) I_{MiCa} conducted Ca^{2+} and Sr^{2+} equally, whereas Mg^{2+} was not measurably permeant (Fig. 3a). The relative divalent ion conductance was $\text{Ca}^{2+} \approx \text{Sr}^{2+} \gg \text{Mn}^{2+} \approx \text{Ba}^{2+}$ (Fig. 3b), identical to the relative divalent ion permeability of the MCU^{13,20–22}. I_{MiCa} currents carried by any of these four permeant divalent ions were substantially reduced by low concentrations of other permeant divalent ions, indicating competition for binding at the selectivity site. The monovalent ions K^+ and Na^+ do not contribute to I_{MiCa} in the presence of $[\text{Ca}^{2+}]_c$ (Fig. 1d; see also Supplementary Fig. 1). We refer to MiCa (mitochondrial calcium) as the highly Ca^{2+} -selective conductance producing large current

densities in mitochondria, consistent with an ion channel.

Ion channel Ca^{2+} selectivity is the result of high-affinity Ca^{2+} binding to the permeation pathway, and removal of external Ca^{2+} enables these ion channels to conduct monovalent ions^{23–26}. Indeed, in low-divalent (5 mM EGTA/5 mM EDTA) solution, we recorded a RuR- and Ru360-sensitive linear Na^+ current (I_{MiNa}). Interestingly, I_{MiNa} was relatively impermeant to K^+ (Fig. 3c, top row). Increasing $[\text{Ca}^{2+}]_c$ to 10 nM partially restored Ca^{2+} selectivity, whereas 100 nM led to complete inhibition of the Na^+ conductance (Fig. 3c, bottom left panel; estimated Ca^{2+} -binding equilibrium constant ≤ 2 nM; see Methods). Although MiCa does not conduct Mg^{2+} , in the absence of other divalent ions MiCa (I_{MiNa}) binds Mg^{2+} (Fig. 3c, bottom right panel) with an estimated binding constant ≤ 250 nM (Methods). In 0.2 mM $[\text{Ca}^{2+}]_c$, 0.5 mM $[\text{Mg}^{2+}]_c$ reduced I_{MiCa} to $41 \pm 8\%$ of its conductance in nominally Mg^{2+} -free solution (not shown). Thus, I_{MiCa} will be reduced by local cytosolic $[\text{Mg}^{2+}]_c$, in agreement with data on the MCU^{27,28}. On the basis of its high sensitivity to RuR, Ru360 and Ca^{2+} , we conclude that I_{MiNa} is the monovalent current in divalent-free conditions.

Single MiCa channel activity was recorded from inside-out mitoplast patches. Under the ionic conditions used for the whole-mitoplast configuration, 3–7 channels were active in each patch and they closely mimicked the inward-rectifying I_{MiCa} current–voltage

relationship (Fig. 4a). This density of single channels agrees with the high I_{MiCa} current density in whole mitoplasts. Single-channel current amplitudes in 105 mM $[\text{Sr}^{2+}]_c$ (pipette) conditions were similar to those in 105 mM $[\text{Ca}^{2+}]_c$ conditions; however, in agreement with whole-mitoplast I_{MiCa} selectivity, 105 mM $[\text{Mg}^{2+}]_c$ currents were not observed. RuR or Ru360 (200 nM; pipette) rapidly blocked inward single-channel activity on stepping to -180 mV (Fig. 4e). Block by 200 nM RuR and Ru360 was relieved on returning to potentials above 0 mV (not shown). Thus, RuR and Ru360 were ineffective in blocking i_{MiCa} outward current (where i indicates unitary current).

In symmetrical 105 mM $[\text{Ca}^{2+}]_c$ conditions, outward single-channel openings at positive voltages (always correlated with inward openings at negative voltages) rapidly flickered between open and closed states (Fig. 4a). Substitution of bath (matrix) Ca^{2+} by Mg^{2+} abolished these outward channels (Fig. 4b), whereas Sr^{2+} -mediated outward currents were comparable to those in Ca^{2+} . In symmetrical 150 mM Na-gluconate low-divalent solution, single i_{MiNa} Na^+ channel currents had about tenfold higher conductance than i_{MiCa} , were selective for Na^+ over K^+ , exhibited multiple conductance states, and were blocked by 200 nM RuR. The rapid flickering of outward single-channel currents observed in 105 mM Ca^{2+} were not observed for the outward monovalent current. This suggests that open channels are transiently blocked by outward-going Ca^{2+}

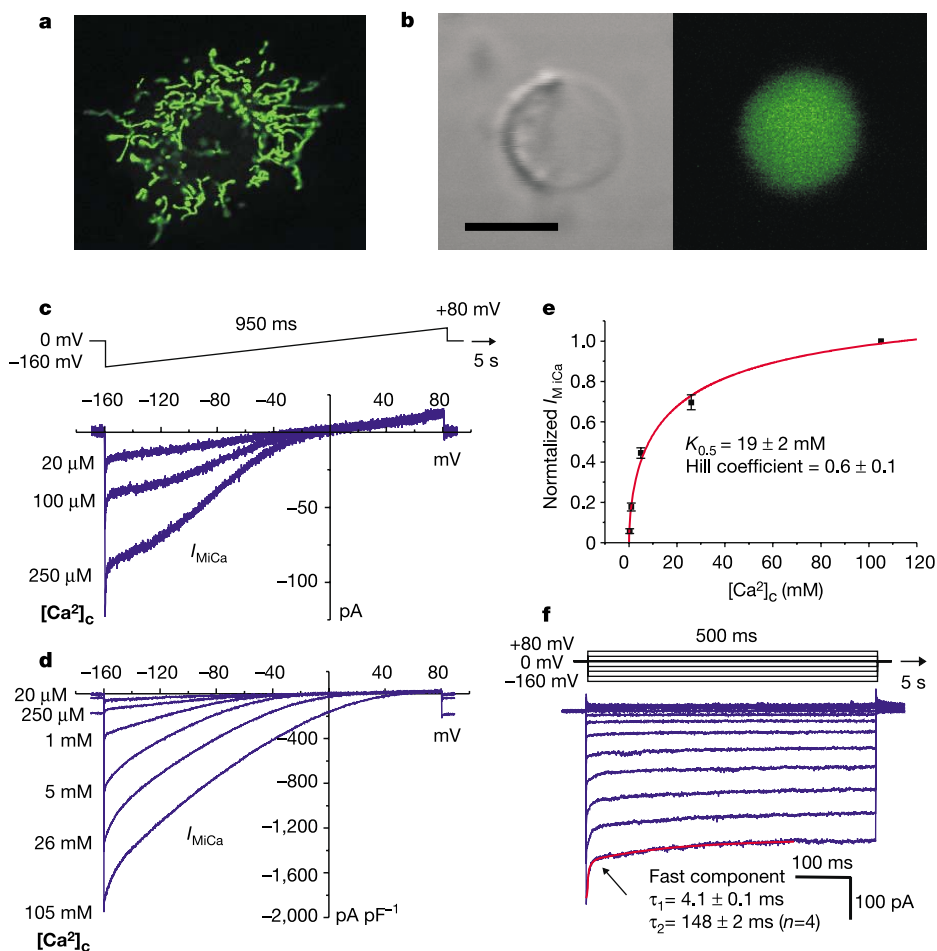


Figure 1 Ca^{2+} current through the inner mitochondrial membrane. **a**, COS-7 cell transfected with mitochondrial-targeted YFP. **b**, Transmitted and fluorescent images of a mito-YFP-labelled mitoplast. Scale bar: 3 μm . **c**, I_{MiCa} elicited by voltage ramps. Negative current flows from the cytoplasmic face of the inner mitochondrial membrane (bath) to the interior of the mitoplast. Bath = 0 mV. Inside mitoplast (pipette), Cs-gluconate solution;

bath, HEPES-Tris solution. **d**, I_{MiCa} increases with bath $[\text{Ca}^{2+}]_c$. Bath, Na-gluconate solution. **e**, Peak amplitude of I_{MiCa} at -160 mV at varying $[\text{Ca}^{2+}]_c$. Currents were normalized to peak I_{MiCa} in 105 mM $[\text{Ca}^{2+}]_c$ and fitted by $I = I_{\text{max}} / (1 + (K_{0.5} / [\text{Ca}^{2+}])^n)$; $n = 5$. **f**, I_{MiCa} in response to voltage steps (5 mM $[\text{Ca}^{2+}]_c$ Na-gluconate solution); $\Delta V = 20$ mV.

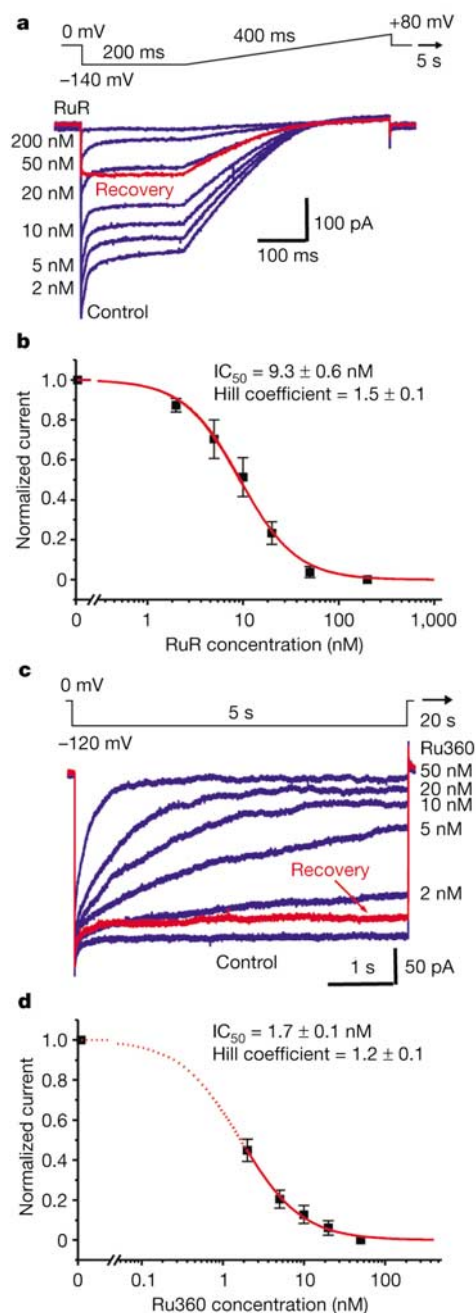


Figure 2 RuR and Ru360 sensitivity. **a**, RuR (2–200 nM) inhibited I_{MiCa} which did not fully recover by 7 min after washout (red trace). The pipette (intramitoplast) contained Na-gluconate solution whereas the HEPES-Tris bath solution contained 5 mM Ca^{2+} . **b**, Dose response of I_{MiCa} to RuR ($n = 5$). **c**, I_{MiCa} inhibition by the MCU blocker Ru360. I_{MiCa} recovered 4 min after washout (red trace). The pipette was filled with Na-gluconate solution (no EGTA/EDTA). **d**, Estimated dose–response curve of I_{MiCa} in response to Ru360. The time courses of inhibition were fitted by single exponential functions with $\tau = 5 \pm 2$ s ($n = 4$) at 2 nM and 270 ± 160 ms ($n = 4$) at 50 nM. Steady state inhibition is estimated from fits for 2 and 5 nM.

ions. Indeed, the amplitude of ensemble-averaged Ca^{2+} currents from multi-channel patches was about nine times less in the outward than in the inward direction (Fig. 4c).

To determine single-channel kinetics from patches containing only one active channel (Fig. 4d), smaller areas of membrane were patched by using greatly reduced pipette tip diameters. In symmetrical 105 mM CaCl_2 solution, the activity of single Ca^{2+}

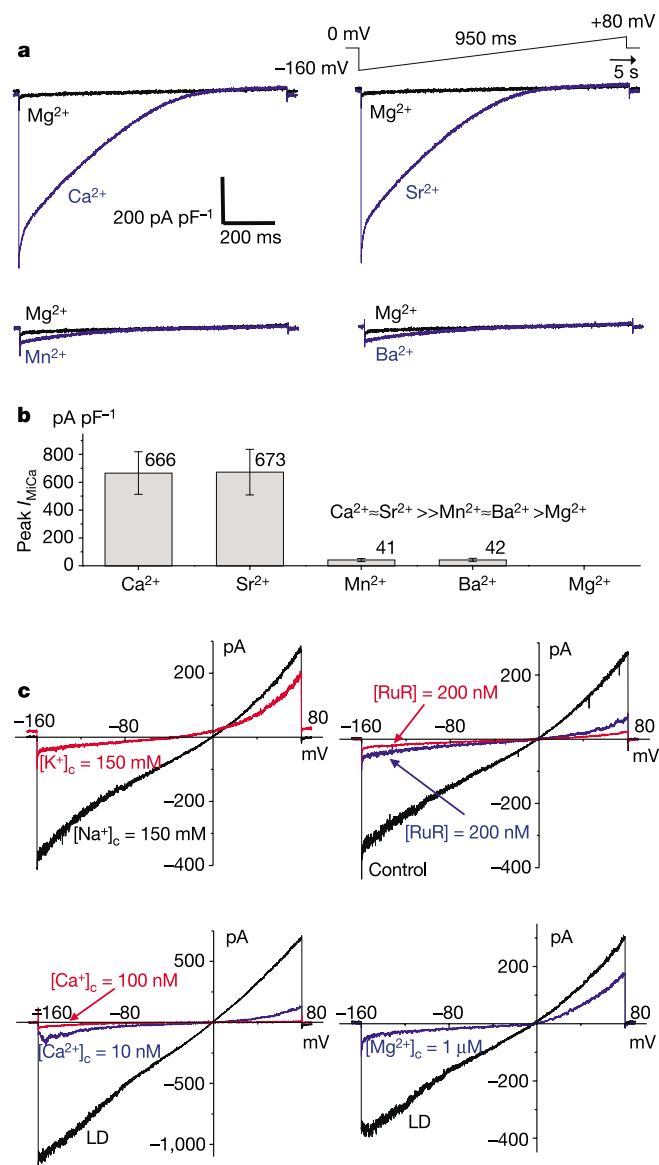


Figure 3 I_{MiCa} selectivity. **a**, I_{MiCa} in 5 mM Ca^{2+} , Sr^{2+} , Mn^{2+} , Ba^{2+} (blue traces) or Mg^{2+} (black traces). Recordings are from the same mitoplast in HEPES-Tris-based solution, washed between applications with 1 mM EGTA/EDTA solution. Pipettes contained Na-gluconate solution. **b**, Relative conductance of I_{MiCa} to divalent cations. Averaged peak current densities at -160 mV ($n = 5$; $\pm$ s.d.). **c**, I_{MiCa} conducts Na^{+} in the absence of divalent ions (I_{MiNa}). Top left panel: I_{MiNa} in Na-gluconate (black trace) and K-gluconate (red trace) low-divalent (LD) solution. Top right panel: I_{MiNa} Na^{+} (LD) current was inhibited by RuR and Ru360. Bottom left panel: $[\text{Ca}^{2+}]_c$ at 10 nM and 100 nM inhibited I_{MiNa} . Bottom right panel: $[\text{Mg}^{2+}]_c$ at 1 μM inhibited I_{MiNa} .

currents was time-independent at all voltages between -200 and $+140$ mV. Single MiCa channels had multiple subconductance states between 2.6 pS and 5.2 pS (-160 mV). Most importantly, the open probability (P_{open}) of the channel was about 99% at -200 mV and declined to approximately 11% at -80 mV. This dependence of P_{open} on membrane potential accounts for much of the inward rectification of the macroscopic I_{MiCa} current over this range of potentials (Supplementary Fig. 2).

In summary, a unique Ca^{2+} -selective channel of the inner mitochondrial membrane accounts for the properties of the MCU. MiCa is the only known functional intracellular Ca^{2+} -selective channel (inositol-1,4,5-trisphosphate and ryanodine

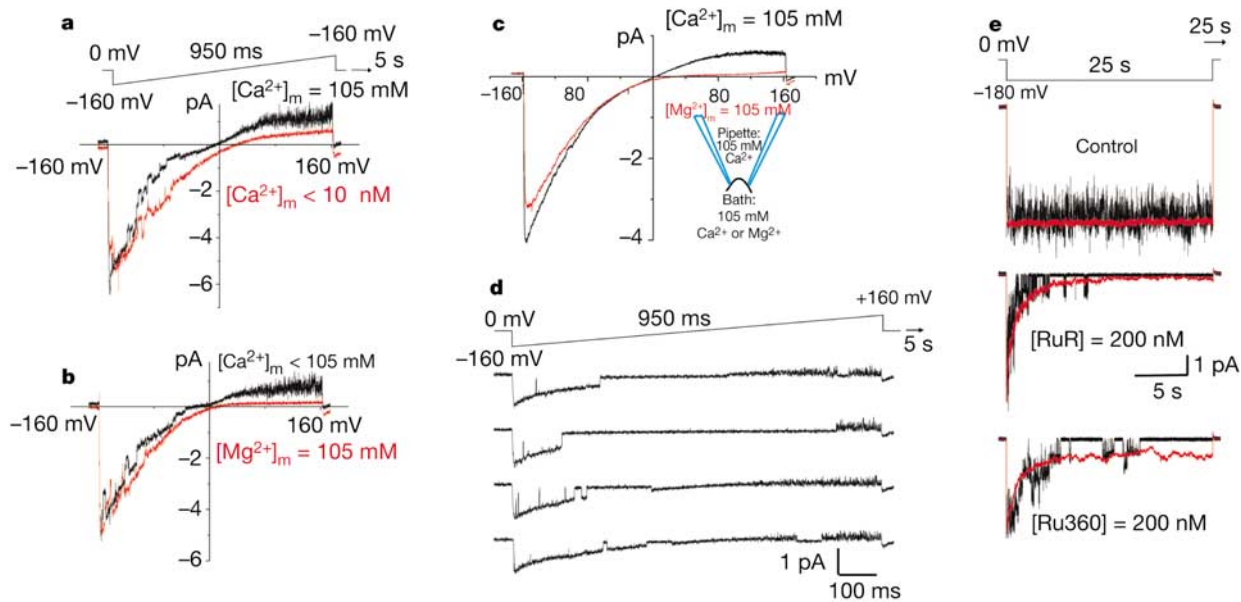


Figure 4 Single i_{MiCa} Ca^{2+} channels from inner mitochondrial membrane inside-out patches (105 mM $CaCl_2$ solution at the cytoplasmic surface (pipette)). **a**, Multi-channel patch currents. The bath contained 150 mM Na-gluconate, 1 mM EGTA/EDTA (free Ca^{2+} concentration at the matrix surface of the inner membrane ($[Ca^{2+}]_m$) < 10 nM) (red trace) or 105 mM $[Ca^{2+}]_m$ (black trace). **b**, i_{MiCa} in 105 mM $[Ca^{2+}]_m$ (black trace) or 105 mM

$[Mg^{2+}]_m$ (red trace). **c**, Ensemble averages of 45 multi-channel responses. Voltage ramp: -160 mV to +160 mV. **d**, Rare single i_{MiCa} . $[Ca^{2+}]_m$ and $[Ca^{2+}]_c$ were both 105 mM. Note the fast open/close kinetics of the attenuated outward current. **e**, [RuR]_c or [Ru360]_c in the pipette inhibited i_{MiCa} at 200 nM. Red traces indicate ensemble averages of 20 multi-channel responses.

receptors are nonselective, but are Ca^{2+} -permeant). MiCa is Ca^{2+} -selective even in low cytoplasmic $[Ca^{2+}]$, it conducts Ca^{2+} preferentially into mitochondria, has relative divalent conductances $Ca^{2+} \approx Sr^{2+} \gg Mn^{2+} \approx Ba^{2+}$, is blocked by nanomolar concentrations of RuR and Ru360, and can provide sufficient current densities to explain mitochondrial uptake by the MCU.

There is one marked difference between our data and Ca^{2+} flux studies using suspensions of isolated mitochondria. Here, in 105 mM $[Ca^{2+}]_c$, I_{MiCa} is close to saturation with a maximum ion flux through the single MiCa molecule of approximately 5×10^6 Ca^{2+} s^{-1} (half-activation constant, $K_{0.5} = 19$ mM), significantly higher than the approximately 2×10^4 Ca^{2+} s^{-1} and $K_{0.5}$ of about 10 μ M estimated for MCU from suspensions of isolated mitochondria^{11,13}. This difference in flux estimates stems from the fact that in mitochondrial suspensions, the rapid dissipation of the mitochondrial potential during Ca^{2+} entry^{11,12} would appear to cause saturation of Ca^{2+} flux at much lower $[Ca^{2+}]_c$. At $[Ca^{2+}]_c$ above 100 μ M, I_{MiCa} current densities are so large that the potential across the inner membrane can only be maintained by voltage clamp. On the basis of an I_{MiCa} density of 1,000–2,000 pA pF^{-1} , $> 90\%$ open probability, and 0.5–1 pA single-channel amplitude at $[Ca^{2+}]_c = 105$ mM and $V_m = -160$ mV, MiCa channel density in the inner mitochondrial membrane is approximately 10–40 channels per μm^2 , slightly lower than voltage-gated Ca^{2+} channel density (about 100 per μm^2) (ref. 29) in excitable cell membranes. Nonetheless, at micromolar concentrations of $[Ca^{2+}]_c$, MiCa channels provide current densities comparable to those of voltage-gated Ca^{2+} channels at millimolar Ca^{2+} concentrations. This feat is accomplished by the large electrochemical gradient for Ca^{2+} across the inner membrane of energized mitochondria and the unusually high probability of MiCa being in the open state. Ca^{2+} -dependent inactivation, a negative feedback mechanism for Ca^{2+} -permeant channels²⁹, is surprisingly absent for MiCa at normal potentials for energized mitochondria.

Na^+ current through MiCa appears only at extremely low concentrations of Ca^{2+} , suggesting that binding sites inside the

pore have an exceptionally high affinity (dissociation constant ≤ 2 nM) for Ca^{2+} . This property guarantees its high selectivity for cytosolic Ca^{2+} where monovalent cations outnumber Ca^{2+} ions by 10^3 – 10^6 -fold. MiCa is impermeant to Mg^{2+} and K^+ , thus preventing mitochondrial depolarization by these abundant ions. In short the high Ca^{2+} selectivity of MiCa results in mitochondrial Ca^{2+} accumulation with the least dissipation of mitochondrial potential. On the basis of the properties demonstrated in our experiments, we conclude that the mitochondrial Ca^{2+} uniporter is the Ca^{2+} -selective ion channel MiCa. □

Methods

Preparation of the mitoplasts

COS-7 cells were homogenized by a Teflon pestle in a glass grinder with 250 mM sucrose, 5 mM HEPES and 1 mM EGTA (pH 7.2 with KOH). Mitochondria were isolated from the lysed COS-7 cell by differential centrifugation in the same solution³⁰. To obtain mitoplasts, mitochondria were subjected to osmotic shock for 5 min in hypotonic solution containing 5 mM sucrose, 5 mM HEPES and 1 mM EGTA (pH 7.2 with KOH). Mitoplasts were sedimented from this solution by centrifugation at 3,920g_{max} for 5 min and subsequently re-suspended in 750 mM KCl, 100 mM HEPES and 1 mM EGTA (pH 7.2 with KOH). Mitoplast isolation was carried out at 2°C and mitoplasts stored on ice. Just before experiments, 1–3 μ l of the mitoplast suspension was added to the experimental solution containing 150 mM Na-gluconate, 10 mM HEPES, 1 mM EGTA and 1 mM EDTA (pH 7.2 with NaOH). Isolated mitoplasts were transparent vesicles with one to several black spots, presumably formed by the remnants of the outer membrane in contact sites between two mitochondrial membranes. In some experiments, mitoplasts were isolated from COS-7 cells transfected with yellow fluorescent protein (YFP) targeted to the inner membrane of mitochondria (BD Clontech) to confirm the mitochondrial origin of the vesicles. Mitoplasts of 2–5 μ m diameter (membrane capacitance, $C_m = 0.35$ –1 pF) were used for patch-clamp experiments.

Whole-mitoplast recordings

After formation of a GΩ seal between the patch-clamp pipette and inner mitochondrial membrane, capacitance transients were completely compensated. Voltage steps of 200–500 mV and 3–300 ms were then applied to rupture the membrane and obtain the whole-mitoplast configuration as monitored by reappearance of capacitance transients and increase in baseline noise. In some experiments BSA (0.05%) was added to the bath solution to increase the survival of mitoplasts. Elevation of pipette solution tonicity ($\times 1.5$ bath solution) significantly reduced the access resistance associated with the mitoplast shrinkage after break-in. The ionic compositions of pipette and bath solutions were chosen to reduce potential contamination by K^+ and Cl^- currents, especially in the inward

direction. Three different pipette solutions were used: Cs-gluconate solution (150 mM CsOH, 5 mM CsCl, 135 mM sucrose, 10 mM HEPES, 1.5 mM EGTA and 1.5 mM EDTA (pH 7.2 with D-gluconic acid)); Na-gluconate solution (150 mM Na-gluconate, 5 mM NaCl, 135 mM sucrose, 10 mM HEPES, 1.5 mM EGTA and 1.5 mM EDTA (pH 7.2 with NaOH)); Na-gluconate solution without Ca^{2+} buffer (150 mM Na-gluconate, 5 mM NaCl, 135 mM sucrose, 10 mM HEPES (pH 7.2 with NaOH)).

Signals were recorded using an Axopatch 200B patch-clamp amplifier, filtered at 5 kHz, and sampled at 10 kHz. Capacitance of the whole-mitoplast membrane ranged from 0.35 to 1 pF. Access resistances, usually 30–60 M Ω , were compensated by about 70%. Bath Ca^{2+} concentration was varied from 20 μM to 5 mM by addition of the corresponding amount of 1 M CaCl_2 stock solution into one of two solutions: Na-gluconate solution (150 mM Na-gluconate, 10 mM HEPES (pH 7.2 with NaOH)) or HEPES-Tris solution (205 mM HEPES (pH 7.2 with Tris base)). Bath solution with 105 mM $[\text{Ca}^{2+}]_c$ (105 mM CaCl_2 solution) contained 105 mM CaCl_2 , 10 mM HEPES (pH 7.2 with Tris base or NaOH). This solution was dissolved with Na-gluconate solution or HEPES-Tris solution to obtain 26 mM $[\text{Ca}^{2+}]_c$. Bath solutions with 5 mM of Sr^{2+} , Ba^{2+} , Mn^{2+} or Mg^{2+} were obtained by addition of 1 M stock solution of the chloride salt into HEPES-Tris solution. HEPES-K bath solution (KOH 75 mM (pH 7.2 with about 210 mM HEPES)); HEPES-Na bath solution (NaOH 75 mM (pH 7.2 with about 210 mM HEPES)); Na-gluconate and K-gluconate low-divalent solutions (150 mM Na(K)-gluconate, 10 mM HEPES, 5 mM EGTA, 5 mM EDTA (pH 7.2 with NaOH)). CaCl_2 or MgCl_2 was added to Na-gluconate low-divalent (divalent-free) solution in the amounts calculated by the WinMAXC v2.05 program (C. Patton, Stanford University) to obtain free Ca^{2+} and Mg^{2+} concentrations. Osmolarity was approximately 280 mmol kg $^{-1}$. Statistical data was calculated as the mean $\pm$ s.d.

Single-channel recordings

All single-channel recordings were made in the inside-out configuration of the patch-clamp technique. Patches were excised from the inner mitochondrial membrane in a bath solution containing 150 mM Na-gluconate, 10 mM HEPES, 1 mM EGTA and 1 mM EDTA (pH 7.2 with NaOH). Pipettes were filled with 105 mM CaCl_2 solution. A total of 105 mM CaCl_2 solution was also used as the bath solution for single channels and whole-mitoplast recordings. Signals were filtered at 1 kHz and sampled at 5 kHz. In Supplementary Fig. 2, traces were filtered at 200 Hz before amplitude analysis.

Estimation of Ca^{2+} - and Mg^{2+} -binding constants

Ca^{2+} buffering is unreliable below 10 nM. To estimate the MiNa Ca^{2+} -binding constant, we assumed that I_{MiNa} in low-divalent solution was $\leq I_{\text{MiNa}}$ at theoretical 0 $[\text{Ca}^{2+}]_c$. Given that I_{MiNa} at 10 nM $[\text{Ca}^{2+}]_c$ was about eightfold less than I_{MiNa} in low-divalent solution, the Ca^{2+} -binding equilibrium constant calculated from the Hill equation was ≤ 2 nM assuming a Hill coefficient of 1. Similarly, the MiNa -binding constant for Mg^{2+} was estimated to be ≤ 250 nM.

Received 20 August; accepted 18 November 2003; doi:10.1038/nature02246.

- Rizzuto, R., Bernardi, P. & Pozzan, T. Mitochondria as all-round players of the calcium game. *J. Physiol. (Lond.)* **529**, 37–47 (2000).
- Montero, M. et al. Chromaffin-cell stimulation triggers fast millimolar mitochondrial Ca^{2+} transients that modulate secretion. *Nature Cell Biol.* **2**, 57–61 (2000).
- McCormack, J. G., Halestrap, A. P. & Denton, R. M. Role of calcium ions in regulation of mammalian intramitochondrial metabolism. *Physiol. Rev.* **70**, 391–425 (1990).
- Hajnóczky, G., Robb-Gaspers, L. D., Seitz, M. B. & Thomas, A. P. Decoding of cytosolic calcium oscillations in the mitochondria. *Cell* **82**, 415–424 (1995).
- Jouaville, L. S., Iachas, F., Holmuhamedov, E. L., Camacho, P. & Lechleiter, J. D. Synchronization of calcium waves by mitochondrial substrates in *Xenopus laevis* oocytes. *Nature* **377**, 438–441 (1995).
- Budd, S. L. & Nicholls, D. G. A re-evaluation of the role of mitochondria in neuronal Ca^{2+} homeostasis. *J. Neurochem.* **66**, 403–411 (1996).
- Babcock, D. F., Herrington, J., Goodwin, P. C., Park, Y. B. & Hille, B. Mitochondrial participation in the intracellular Ca^{2+} network. *J. Cell Biol.* **136**, 833–844 (1997).
- Boitier, E., Rea, R. & Duchen, M. R. Mitochondria exert a negative feedback on the propagation of intracellular Ca^{2+} waves in rat cortical astrocytes. *J. Cell Biol.* **145**, 795–808 (1999).
- Scorrano, L. et al. BAX and BAK regulation of endoplasmic reticulum Ca^{2+} : a control point for apoptosis. *Science* **300**, 135–139 (2003).
- Demaurex, N. & Distelhorst, C. Cell biology. Apoptosis—the calcium connection. *Science* **300**, 65–67 (2003).
- Gunter, K. K. & Gunter, T. E. Transport of calcium by mitochondria. *J. Bioenerg. Biomembr.* **26**, 471–485 (1994).
- Bernardi, P. Mitochondrial transport of cations: channels, exchangers, and permeability transition. *Physiol. Rev.* **79**, 1127–1155 (1999).
- Gunter, T. E. & Pfeiffer, D. R. Mechanisms by which mitochondria transport calcium. *Am. J. Physiol.* **258**, C755–C786 (1990).
- DeLuca, H. F. & Engstrom, G. W. Calcium uptake by rat kidney mitochondria. *Proc. Natl Acad. Sci. USA* **47**, 1744–1750 (1961).
- Rizzuto, R., Brini, M., Murgia, M. & Pozzan, T. Microdomains with high Ca^{2+} close to IP_3 -sensitive channels that are sensed by neighboring mitochondria. *Science* **262**, 744–747 (1993).
- Neher, E. Vesicle pools and Ca^{2+} microdomains: new tools for understanding their roles in neurotransmitter release. *Neuron* **20**, 389–399 (1998).
- Tashmukhamedov, B. A., Gagelgans, A. I., Mamatkulov, K. & Makhmudova, E. M. Inhibition of Ca^{2+} transport in mitochondria by selective blockade of membrane mucopolysaccharides by hexamine cobalt chloride. *FEBS Lett.* **28**, 239–242 (1972).
- Ying, W. L., Emerson, J., Clarke, M. J. & Sanadi, D. R. Inhibition of mitochondrial calcium ion transport by an oxo-bridged dinuclear ruthenium amine complex. *Biochemistry* **30**, 4949–4952 (1991).
- Matlib, M. A. et al. Oxygen-bridged dinuclear ruthenium amine complex specifically inhibits Ca^{2+}

uptake into mitochondria *in vitro* and *in situ* in single cardiac myocytes. *J. Biol. Chem.* **273**, 10223–10231 (1998).

- Drahota, Z., Gazzotti, P., Carafoli, E. & Rossi, C. S. A comparison of the effects of different divalent cations on a number of mitochondrial reactions linked to ion translocation. *Arch. Biochem. Biophys.* **130**, 267–273 (1969).
- Vainio, H., Mela, L. & Chance, B. Energy dependent bivalent cation translocation in rat liver mitochondria. *Eur. J. Biochem.* **12**, 387–391 (1970).
- Selwyn, M. J., Dawson, A. P. & Dunnett, S. J. Calcium transport in mitochondria. *FEBS Lett.* **10**, 1–5 (1970).
- Kostyuk, P. G. & Krishtal, O. A. Effects of calcium and calcium-chelating agents on the inward and outward current in the membrane of mollusc neurones. *J. Physiol. (Lond.)* **270**, 569–580 (1977).
- Almers, W., McCleskey, E. W. & Palade, P. T. A non-selective cation conductance in frog muscle membrane blocked by micromolar external calcium ions. *J. Physiol. (Lond.)* **353**, 565–583 (1984).
- McCleskey, E. W. & Almers, W. The Ca channel in skeletal muscle is a large pore. *Proc. Natl Acad. Sci. USA* **82**, 7149–7153 (1985).
- Prakriya, M. & Lewis, R. S. CRAC channels: activation, permeation, and the search for a molecular identity. *Cell Calcium* **33**, 311–321 (2003).
- Vinogradov, A. & Scarpa, A. The initial velocities of calcium uptake by rat liver mitochondria. *J. Biol. Chem.* **248**, 5527–5531 (1973).
- Bragadin, M., Pozzan, T. & Azzzone, G. F. Kinetics of Ca^{2+} carrier in rat liver mitochondria. *Biochemistry* **18**, 5972–5978 (1979).
- Hille, B. *Ion Channels of Excitable Membranes* (Sinauer Associates, Sunderland, Massachusetts, 2001).
- Cannon, B. & Lindberg, O. Mitochondria from brown adipose tissue: isolation and properties. *Methods Enzymol.* **55**, 65–78 (1979).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank J. Borecky for advice on the whole-mitoplast configuration; L. DeFelice, H. Xu, B. Desai, V. Sandler and P. Smith for discussions; and S. Gapon and Y. Manasian for technical assistance.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.E.C. (dclapham@enders.tch.harvard.edu).

Two mitotic kinesins cooperate to drive sister chromatid separation during anaphase

Gregory C. Rogers¹, Stephen L. Rogers², Tamara A. Schwimmer¹, Stephanie C. Ems-McClung³, Claire E. Walczak³, Ronald D. Vale², Jonathan M. Scholey⁴ & David J. Sharp¹

¹Department of Physiology and Biophysics, Albert Einstein College of Medicine, Bronx, New York 10461, USA

²The Howard Hughes Medical Institute and Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143, USA

³Medical Sciences Program, Indiana University, Bloomington, Indiana 47405, USA

⁴Center for Genetics and Development and Section of Molecular and Cellular Biology, University of California, Davis, California 95616, USA

During anaphase identical sister chromatids separate and move towards opposite poles of the mitotic spindle^{1,2}. In the spindle, kinetochore microtubules³ have their plus ends embedded in the kinetochore and their minus ends at the spindle pole. Two models have been proposed to account for the movement of chromatids during anaphase. In the ‘Pac-Man’ model, kinetochores induce the depolymerization of kinetochore microtubules at their plus ends, which allows chromatids to move towards the pole by ‘chewing up’ microtubule tracks^{4,5}. In the ‘poleward flux’ model, kinetochores anchor kinetochore microtubules and chromatids are pulled towards the poles through the depolymerization of kinetochore microtubules at the minus ends⁶. Here, we show that two functionally distinct microtubule-destabilizing KifI kinesin

enzymes (so named because they possess a kinesin-like ATPase domain positioned internally within the polypeptide) are responsible for normal chromatid-to-pole motion in *Drosophila*. One of them, KLP59C, is required to depolymerize kinetochore microtubules at their kinetochore-associated plus ends, thereby contributing to chromatid motility through a Pac-Man-based mechanism. The other, KLP10A, is required to depolymerize microtubules at their pole-associated minus ends, thereby moving chromatids by means of poleward flux.

Although molecules involved in either Pac-Man- or poleward-flux-based motility mechanisms (see Supplementary Fig. S1) have not been identified previously, members of the KinI subfamily of kinesins are logical candidates for both⁷. These proteins bind directly to the ends of microtubules and promote their disassembly *in vitro*^{8–10}. Moreover, the inhibition of KinI proteins in mitotic cells from various systems has been shown to disrupt mitosis^{11–13}. To identify novel KinI family members in *Drosophila melanogaster*, we searched *Drosophila* genome databases for genes that share significant identity with the known KinI catalytic domains from other species. This revealed three different *D. melanogaster* genes predicted to encode products containing internal kinesin domains highly similar to vertebrate KinI (65% amino acid identity on average; Supplementary Fig. S2). We refer to these as KLP10A, KLP59C and KLP59D (kinesin-like protein at cytological regions 10A, 59C and 59D; ref. 14). Notably, we observed that purified recombinant polypeptides corresponding to the catalytic domains of KLP10A and KLP59C destabilize microtubules in an ATP-dependent fashion (Supplementary Fig. S3). (Similar analysis

of KLP59D has not yet been performed but the high degree of amino acid identity among the catalytic domains of all three *Drosophila* KinI proteins (66% on average) suggests that it will also possess microtubule-destabilizing activity.) Thus, *Drosophila* probably possesses three distinct microtubule-destabilizing KinI kinesins.

Analysis of *Drosophila* S2 tissue culture cells depleted of each of the three KinI proteins (individually and in various combinations) using double-stranded RNA interference (RNAi) suggests that both KLP10A and KLP59C perform important but distinct mitotic functions (Supplementary Figs S4g, S5 and Table S1). KLP10A depletion by RNAi treatment causes a marked perturbation of mitotic spindle architecture (Fig. 1b–e). In contrast, KLP59C RNAi treatment has no noticeable impact on mitotic spindle structure but does significantly elevate the frequency with which chromosome segregation defects are observed (Fig. 1g–i). Immunolocalization of these proteins in mitotic S2 cells also reveals distinctions that may provide insights into their specific mitotic mechanisms of action. Specifically, KLP10A, similarly to vertebrate KinI proteins, localizes to mitotic centrosomes, spindle poles and centromeres through metaphase (Fig. 1j; see also Supplementary Fig. S6). However, the centromeric localization of KLP10A diminishes markedly at the onset of anaphase, leaving the majority of KLP10A immunostaining on the spindle poles (Fig. 1k). KLP59C, on the other hand, appears to be primarily restricted to centromeric regions of chromosomes during both metaphase and anaphase, and no KLP59C immunostaining is detectable on spindle poles throughout mitosis (Fig. 1l, m; see also Supplementary Fig. S6).

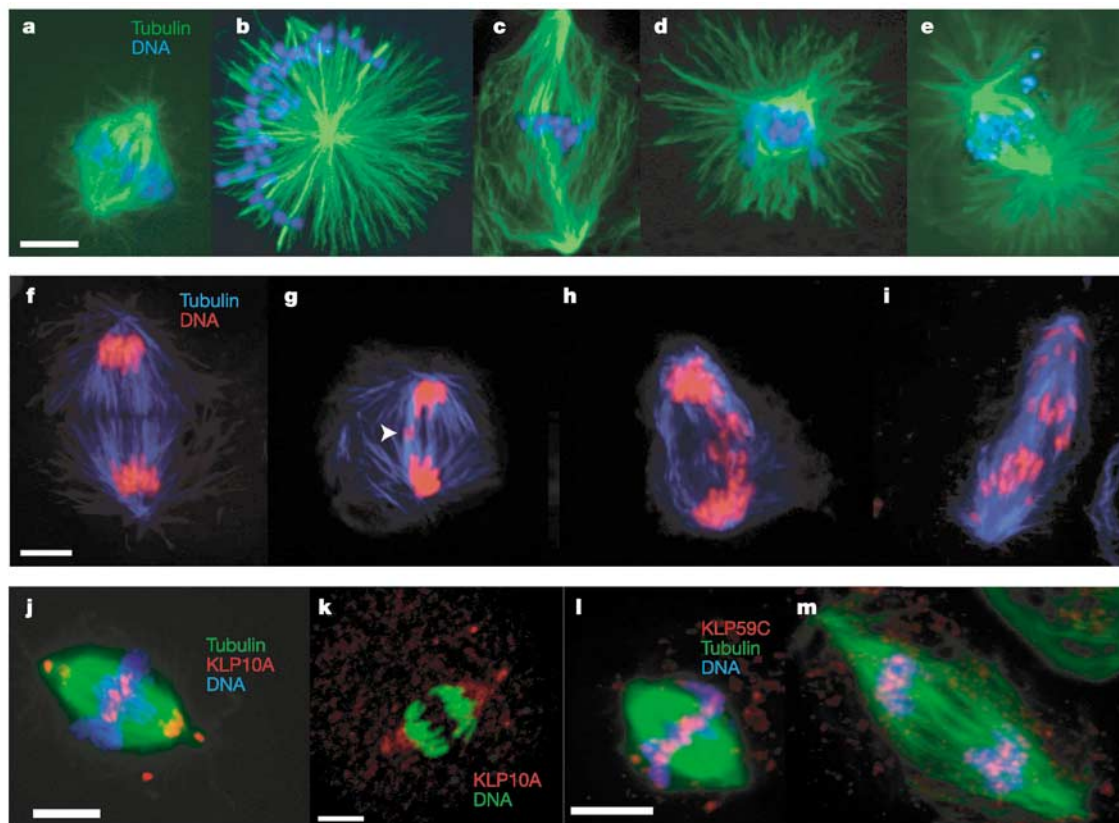


Figure 1 KLP10A and KLP59C perform distinct functions and localize to distinct sites in mitotic cells. **a–e**, Mitotic spindles from control (**a**) and KLP10A RNAi-treated (**b–e**) S2 cultures. Defects observed after KLP10A RNAi treatment consistently fall into one of four classes: monopolar spindles (**b**); long metaphase figures with disorganized central spindles (**c**); microtubule baskets forming around chromosomes (**d**); spindles with

abnormally long astral microtubules (**e**). **f–i**, Anaphase spindles displaying normal (**f**) and defective (**g–i**) chromosome segregation. KLP59C RNAi treatment elevates the frequency of defective chromosome segregation (KLP59C RNAi, $42.9 \pm 1\%$; control, $23 \pm 1\%$; $P < 0.0002$). **j–m**, Localization of KLP10A (**j**, **k**) and KLP59C (**l**, **m**) during metaphase (**j**, **l**) and anaphase (**k**, **m**). Scale bar, 5 mm.

Thus, during chromatid-to-pole motion in anaphase, KLP10A and KLP59C are positioned appropriately to interact with the opposite ends of kinetochore microtubules: KLP10A concentrates around the minus ends of kinetochore microtubules focused at the poles, whereas KLP59C is positioned to act on the plus ends of kinetochore microtubules embedded in the kinetochore. Finally, our analyses of KLP59D do not support a role for this protein in spindle assembly or chromosome segregation, and its cellular functions will be reported elsewhere.

The specific mechanisms of action of mitotic KLP10A and KLP59C are impossible to discern from fixed samples. Therefore, we used living *Drosophila* syncytial blastoderm-stage embryos to elucidate the dynamic events leading to the mitotic phenotypes described in KLP10A- and KLP59C-deficient S2 cultures. These embryos are ideal for such analyses because they contain a monolayer of adjacent nuclei, within the same cytoplasm, that proceed through mitosis synchronously and relatively rapidly¹⁵. Moreover, these cells lack a standard mitotic spindle checkpoint, which normally delays cells in prometaphase in response to spindle damage, allowing mitotic proteins that might function later in the cell cycle to be assessed¹⁶. For our studies, soluble KinI inhibitors (Supplementary Figs S2c, S4a–f and S8a) and fluorescent tubulin or histones were microinjected into living transgenic embryos expressing different green fluorescent protein (GFP)-tagged spindle proteins. This allows the dynamic behaviour of spindle components to be visualized in the presence and absence of specific

KinI inhibitors. (Supplementary Movies 1 and 2 show mitosis in control-injected embryos).

The introduction of KLP10A inhibitors into embryos rapidly alters the organization of mitotic spindle microtubules, as revealed by time-lapse imaging of anti-KLP10A antibody-injected transgenic embryos expressing GFP-tubulin as a microtubule marker (Fig. 2). Within 2 min of antibody injection, the concentration of microtubules at centrosomes increases significantly compared with controls. The relative fluorescent intensities of GFP-tubulin in both prophase and metaphase domains were on average 1.4-fold higher in regions proximal to the injection site compared with mitotic domains at more distal sites (Supplementary Table S2). Thus, KLP10A normally limits the growth/density of microtubules at centrosomes. Subsequently, severe defects in mitotic spindle assembly become apparent (Fig. 2a; see also Supplementary Movie 3). These can be categorized into two classes (Supplementary Table S3). Thirty per cent formed monopolar spindles, which assemble when centrosomes (partially separated during prophase) collapse back together at a rate of $0.019 \mu\text{m s}^{-1}$ after nuclear envelope breakdown (class one). These structures may result from general defects in the organization of centrosomal microtubules. A total of 61% display abnormally long bipolar spindles, which often form adjacent to monopolar spindles (class two). These bipolar spindles elongate continuously during prometaphase through anaphase and attain metaphase lengths approximately twice that of controls (Fig. 2c). Notably, the rate of elongation of these spindles

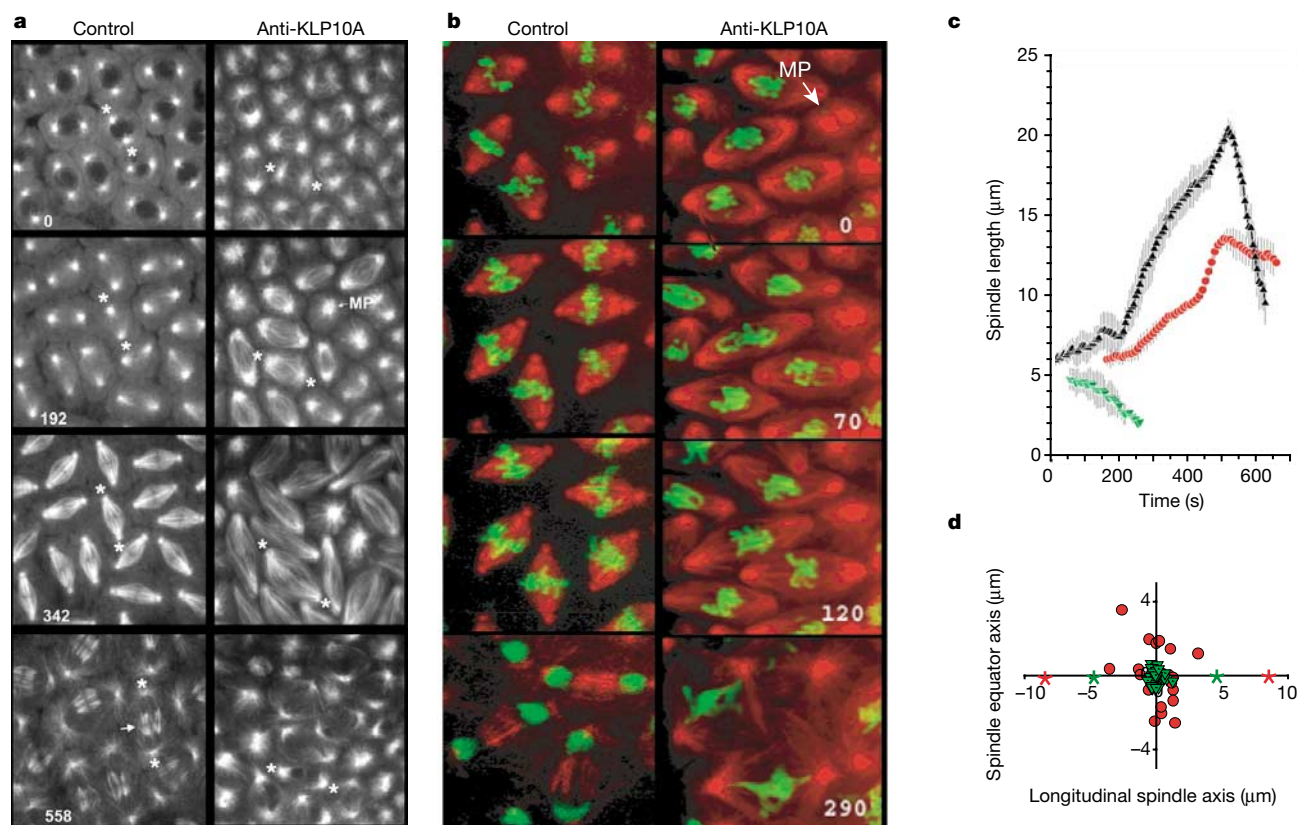


Figure 2 KLP10A inhibition perturbs mitotic spindle assembly and chromosome segregation. **a, b**, Time series of images showing mitosis in control (left) and anti-KLP10A-injected (right) embryos. Embryos in **a** express GFP-tubulin, allowing spindles to be visualized, whereas those in **b** contain rhodamine-tubulin and GFP-histones, allowing simultaneous visualization of spindles (red) and chromosomes (green). Time (in seconds), centrosomes from selected spindles (asterisk), monopolar spindles (MP), and telophase midbody (arrow) are marked. **c**, Plots comparing spindle length over time in control and

anti-KLP10A-injected embryos. Plots begin at prometaphase. Red circles, control injected; green triangles, monopolar spindles from anti-KLP10A-injected embryos; black triangles, bipolar spindles from anti-KLP10A-injected embryos. **d**, Graph of the average centroid positions of GFP-Cid-labelled kinetochores on spindles in control (green triangles) and anti-KLP10A-injected (red circles) embryos. Red stars mark the average position of spindle poles in spindles from KLP10A-inhibited embryos and green stars mark the average position of poles from controls.

before anaphase is nearly identical to the rate of anaphase B spindle elongation in controls (peak rate ($\pm$ s.d.) for anti-KLP10A is $0.061 \pm 0.006 \mu\text{m s}^{-1}$, $n = 10$, and for controls is $0.062 \pm 0.008 \mu\text{m s}^{-1}$, $n = 10$). The implications of this observation are discussed below. Finally, during telophase, midbodies fail to form on these spindles and centrosomes 'snap-back' together.

In addition to spindle defects, KLP10A inhibition in embryos perturbs the movement and segregation of chromosomes (Fig. 2b; see also Supplementary Movie 4). Specifically, the motility of individual chromosomes during prometaphase is often incoherent in treated embryos and subsequently, during metaphase, the majority of chromosomes and kinetochores fail to align properly at the metaphase plate. Instead, they are scattered around the spindle equator (Fig. 2d). Finally, during anaphase, chromatids translocate towards spindle poles at rates significantly slower than controls (average rate: $0.054 \mu\text{m s}^{-1}$ anti-KLP10A; $0.094 \mu\text{m s}^{-1}$ controls; Supplementary Table S4). This results in a significant increase in severe chromosome segregation defects such as stretched or bi-lobed chromosome masses (Supplementary Fig. S7 and Table S3). These anaphase phenotypes are observed regardless of whether kinetochores congressed properly to the metaphase plate or not. Indistinguishable results were obtained using monovalent Fab fragments (Supplementary Fig. S8b) and were corroborated using a recombinant dominant/negative construct (Supplementary Fig. S9).

In contrast to our findings for KLP10A, the introduction of KLP59C inhibitors into embryos has no apparent effect on the organization of mitotic spindles (Fig. 3a, c; see also Supplementary Movie 5 and Table S3). However, KLP59C inhibition does produce significant defects in chromosome positioning and motility (Supplementary Fig. S7 and Table S3). Chromosomes and kinetochores fail to align tightly at the spindle equator during metaphase under these conditions (Fig. 3d). Moreover, KLP59C inhibition prevents

normal chromosome segregation during anaphase (Fig. 3b). Failures in chromosome segregation do not seem to result from decondensation of lagging chromosomes at the metaphase plate, nor are they due to non-disjunction of sister chromatids, because kinetochores do segregate to opposite sides of the chromosomal mass (Fig. 3b). Instead, as with KLP10A, the rate of chromatid-to-pole motion is slowed significantly in KLP59C-inhibited embryos (average rate: $0.042 \mu\text{m s}^{-1}$ anti-KLP59C; $0.094 \mu\text{m s}^{-1}$ controls; Supplementary Table S4). (These effects were reproduced by Fabs and a dominant/negative construct (Supplementary Fig. S8a, c and S9).)

Given the possibility of nonspecific effects, we investigated the mechanism of protein perturbation induced by antibody injections. The antibodies used in our experiments were raised against recombinant amino-terminal targeting domains (NT) of the *Drosophila* KinI kinesins¹¹ (Supplementary Fig. S2c). Thus, we reasoned that injection of these antibodies would specifically inhibit KinI function by displacing endogenous KLP10A and KLP59C from spindles¹³. Fluorescently labelled NT polypeptides mimic the spindle localization of endogenous KLP10A and KLP59C when injected into embryos: the 10A-NT protein localizes to centrosomes, spindle poles and metaphase centromeres, whereas 59C-NT localizes to centromeres (Supplementary Fig. S10). Therefore, these fluorescent constructs could be used to report the behaviour of the endogenous proteins *in vivo*. Consistent with the hypothesis that our KinI antibodies displace endogenous KinI proteins, the 10A-NT and 59C-NT proteins were rapidly mis-localized from spindle and/or chromosomal sites after the injection of anti-KLP10A and anti-KLP59C antibodies, respectively. In contrast, the reciprocal experiments do not produce this effect (for example, anti-KLP59C antibodies do not mis-localize the 10A-NT construct and vice versa). Moreover, injected fluorescently labelled antibodies often form cytoplasmic aggregates that co-localize with aggregates of the

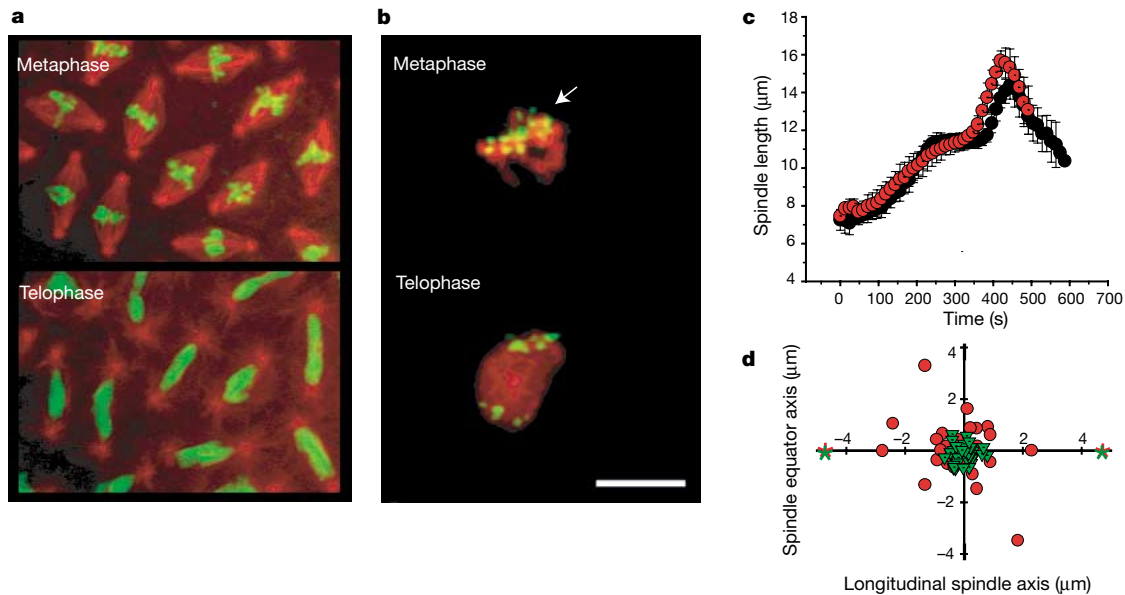


Figure 3 KLP59C inhibition perturbs chromosome segregation but not spindle assembly. **a**, Sequential images showing mitosis (metaphase and telophase) in an anti-KLP59C antibody-injected embryo containing rhodamine-tubulin (red) and GFP-histones (green). **b**, Sequential images illustrating improper kinetochore alignment (metaphase) and failed chromosome segregation (telophase) observed in an embryo containing rhodamine-histones (red) and GFP-Cid-labelled kinetochores (green). Although chromosomes do not segregate into daughter nuclei, sister kinetochores have segregated into two distinct sets at opposite sides of the chromosome mass. The arrow indicates a cluster of

kinetochores that have failed to align properly. **c**, Plots of spindle length over time in control (red circles) and anti-KLP59C antibody-injected (black circles) embryos (beginning at prometaphase). **d**, Graph showing the average centroid positions of kinetochores on spindles from control (green triangles) and anti-KLP59C antibody-injected (red circles) embryos. Red stars mark the average position of spindle poles in spindles from KLP10A-inhibited embryos and green stars mark the average position of poles from controls.

injected NT proteins (Supplementary Fig. S10). Thus, our antibodies have the capacity to specifically precipitate endogenous KLP10A or KLP59C from their sites of action. Finally, it does not appear that the defects observed after KinI inhibition are due to an inhibition of the assembly of other functionally important proteins onto centromeres, kinetochores or centrosomes (where KLP10A and/or KLP59C localize). Indeed, KinI antibody injections do not overtly perturb the localization of a variety of GFP-tagged proteins that normally target to these spindle structures (Figs 2d, 3b, d and 4c; see also Supplementary Fig. S11). Instead, it is most likely that our antibodies specifically prevent targeting of the KinI proteins.

Together, these analyses reveal that both KLP10A and KLP59C perform functions needed to move chromatids poleward during anaphase. Furthermore, KLP10A performs an additional function involved in organizing spindle microtubules.

The capacity of KinI kinesins to depolymerize microtubules *in vitro*¹⁰ (Supplementary Fig. S3), along with our observations that both KLP10A and KLP59C function in chromosome segregation, is consistent with the hypothesis that anaphase chromatid-to-pole motion requires the KinI-induced depolymerization of kinetochore microtubule ends. To test this hypothesis further, we examined the behaviour of spindle microtubules in the presence and absence of specific KLP10A and KLP59C inhibitors. Spindle microtubule dynamics were tracked using fluorescent speckle microscopy (FSM)—a technique in which dilute levels of rhodamine-labelled tubulin are microinjected into cells, resulting in the

random incorporation of fluorescent tubulin ‘speckles’ into the microtubule lattice¹⁷. Speckles serve as markers that allow the direct observation of the dynamic behaviours of individual spindle microtubules. In controls, FSM reveals the persistent poleward movement of tubulin subunits within microtubule polymer lattices; that is, poleward flux (Fig. 4a; see also Supplementary Movie 6), as previously characterized^{18,19}.

The simultaneous imaging of GFP-tagged kinetochores and tubulin speckles provides an *in vivo* method to distinguish between the depolymerization of kinetochore microtubule plus or minus ends; flux requires kinetochore microtubule minus-end depolymerization at spindle poles, and kinetochores that move poleward more rapidly than speckles must depolymerize kinetochore microtubules at their plus ends. Consistent with previous reports, we find that the average rate of poleward flux is identical during metaphase and anaphase A ($0.037 \mu\text{m s}^{-1}$) and is consistently slower than chromatid-to-pole movement (average rate: $0.094 \mu\text{m s}^{-1}$, Fig. 4c; see also Supplementary Table S4)^{18,19}. This suggests that poleward flux can contribute approximately 40% to the rate of anaphase chromatid-to-pole motion whereas Pac-Man-based motility can contribute roughly 60%. Moreover, these data directly demonstrate that kinetochore microtubules simultaneously depolymerize at both plus and minus ends during anaphase in *Drosophila* embryos.

To test the role of KLP10A in spindle microtubule depolymerization, FSM was performed in embryos injected with anti-KLP10A antibodies (Fig. 4a). In contrast to control embryos, fluorescent

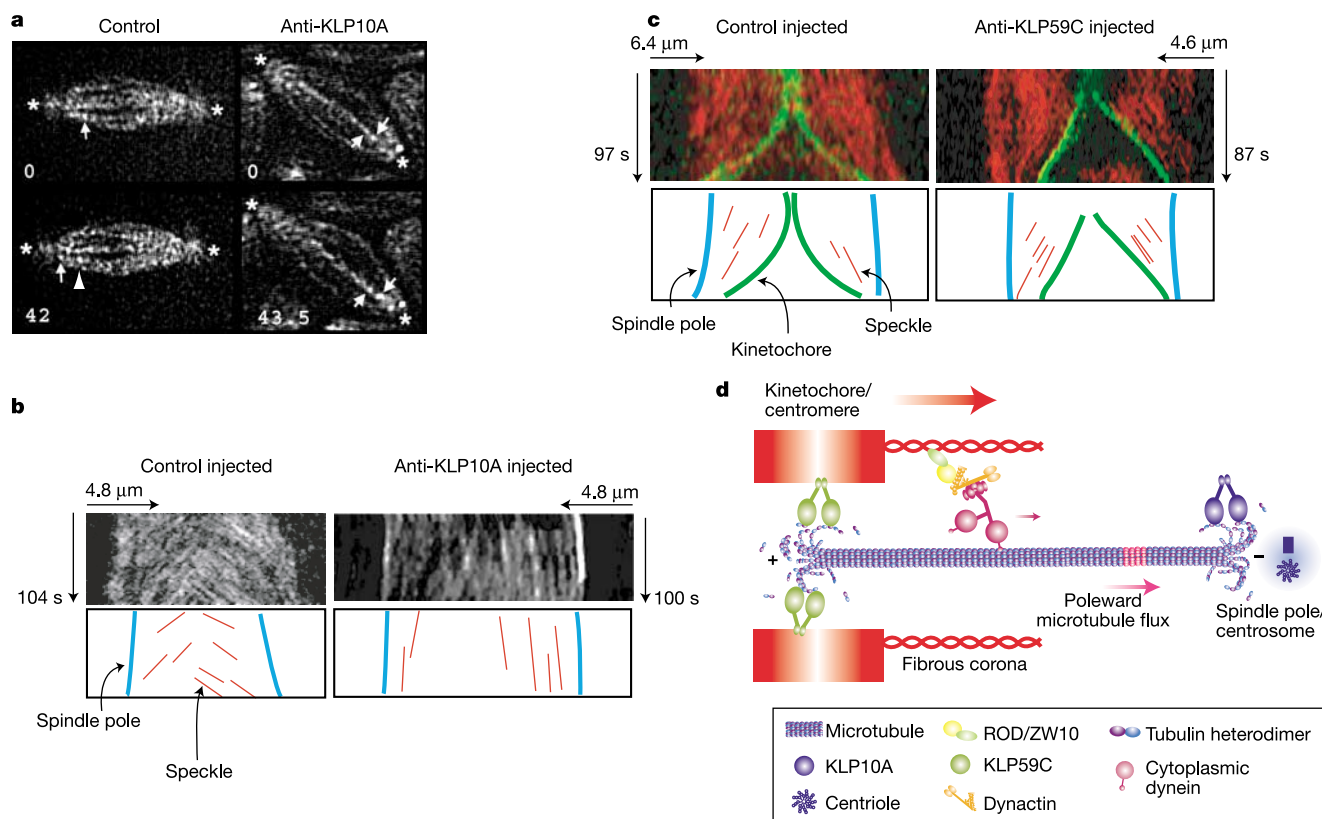


Figure 4 KLP10A is required for poleward flux whereas KLP59C is required for Pac-Man-based chromosome motility. **a**, Sequential FSM images of tubulin-speckle-labelled spindles from control and anti-KLP10A antibody-injected embryos. Time (in seconds) and centrosome positions (asterisk) are indicated. In the left panel, the arrow follows a speckle as it ‘fluxes’ poleward. The arrowhead (bottom) marks the approximate starting position of the speckle. In the right panel, the arrows track the position of two speckles over the same

time period. These show little poleward displacement. **b**, Kymographs (top) and tracings (bottom) of fluorescent tubulin-labelled speckle movement on prometaphase/metaphase spindles from both conditions. **c**, Kymographs and tracings showing the movement of kinetochores (green) and tubulin-labelled speckles (red) on spindles from control and anti-KLP59C antibody-injected embryos. **d**, Model^{9,30}.

tubulin speckles either do not move poleward at detectable velocities or move at velocities that are significantly slowed (Supplementary Movie 7). Kymograph analysis of speckle movements confirms the point tracking of speckles and spindle edges (Fig. 4b; see also Supplementary Table S4). Indeed, >90% of speckles in KLP10A-inhibited spindles either do not have measurable motion or move at a reduced rate (Supplementary Fig. S12). Thus, KLP10A is required for normal poleward flux, making it the first molecular component of this phenomenon identified so far. On the basis of the proximity of this protein to microtubule minus ends at poles, our findings suggest that KLP10A promotes flux by actively depolymerizing kinetochore microtubules at their minus ends.

These data, along with the phenotypes observed after KLP10A inhibition described above, are notable in that they suggest roles for poleward flux in both anaphase chromatid-to-pole motion and in constraining the length of the mitotic spindle through metaphase²⁰ (see above). Although our focus here is on the former finding, the latter is also intriguing given the recent report that poleward flux ceases during anaphase B¹⁸ (which occurs after chromatids have moved completely poleward). Thus, inactivation of spindle-pole-associated KinI proteins may normally serve as a trigger for anaphase spindle elongation.

In contrast to our KLP10A findings, KLP59C inhibition does not reduce the velocity of poleward tubulin speckle movement within spindles (Supplementary Table S4). However, kymographs comparing kinetochore movements to poleward flux reveal that, after KLP59C inhibition, kinetochores move poleward at the rate of flux but no faster, a decrease from control rates by roughly 60% (Fig. 4c; see also Supplementary Table S4). Indeed, under these conditions, kinetochores rarely bypass speckles on poleward-moving spindle microtubules, a phenomenon regularly observed in controls. Thus, KLP59C inhibition causes kinetochores to lose the ability to move along kinetochore microtubules and leaves them to behave solely as kinetochore microtubule anchors. Given the KLP59C localization to centromeres and its ability to destabilize microtubules *in vitro*, we conclude that this protein actively depolymerizes kinetochore microtubule plus ends and thus, it is the first required molecular component of Pac-Man-based chromatid motility identified so far.

Taken together, our findings support the mechanism for anaphase chromatid-to-pole movement shown in Fig. 4d. In this model, poleward driving forces acting on chromosomes result from the simultaneous disassembly of kinetochore microtubules at both minus and plus ends. At the minus ends, KinI-induced disassembly allows chromosome-associated kinetochore microtubules to be driven back into the poles (poleward flux). This activity probably requires additional factors that slide microtubules poleward. At the other end of kinetochore microtubules, centromere-associated KinI proteins act to disassemble kinetochore microtubule plus ends, causing them to shorten (Pac-Man). We propose that the minus-end-directed motor, cytoplasmic dynein, participates in this process by feeding kinetochore microtubules into the kinetochore²¹. It is tempting to speculate that this coupled, KinI-dependent, 'Pac-Man flux' model is conserved among phyla that contain KinI kinesins^{9,11–13,22–25}, although further work is required to establish this. If so, KinI motors will be promising targets for anti-tumour therapy. Although there are probably additional complexities to anaphase chromosome motility, the concerted action of flux and Pac-Man-based motility mechanisms provides the cell with a surprisingly dynamic means of segregating chromosomes. □

Methods

Drosophila KinI sequence data are described in Supplementary Fig. S2 legend.

Double-stranded RNAi

S2 cell culture and RNAi was performed as described²⁶. Molecular methods are described in Supplementary Fig. S4 legend.

Immunofluorescence microscopy

For immunostaining, S2 cells were fixed exactly as described²⁶. Antibodies were diluted to concentrations ranging from 1 to 20 $\mu\text{g ml}^{-1}$ (DM1a (Sigma), 1:200; GTU-88 (Sigma), 1:500; chicken anti-Cid (from G. Karpen), 1:500; rabbit anti-KLP10A 654, 655 and 656, 1:100; rabbit anti-KLP59C 696 and 698, 1:100). Secondary antibodies (Cy2-conjugated, rhodamine red X-conjugated and Cy5-conjugated anti-rabbit, mouse or chicken (Jackson ImmunoResearch)) were used at final dilutions of 1:100. Cells were mounted in Prolong (Molecular Probes). Propidium iodide (Sigma) was used at a concentration of 0.2 mg ml^{-1} . Mitotic S2 chromosomes were purified as described²⁷. Specimens were imaged using a spinning-disk confocal microscope and displayed as maximum intensity projections.

Antibodies

DNA constructs are described in Supplementary Fig. S2 legend. *Escherichia coli*-expressed proteins were purified on either glutathione-Sepharose or amylose resin. Anti-glutathione S transferase (GST)-10A-NT, anti-GST-59C-NT and anti-GST-59D-NT antiserum was produced in rabbits (Covance). Antibodies were affinity-purified against their respective MBP fusion proteins precoupled to Affigel 10/15 (Bio-Rad) and eluted with low-pH buffer.

Microtubule depolymerization assays

Visual and sedimentation analyses of stabilized microtubule depolymerization were performed as previously described¹⁰ with purified baculovirus-expressed (His)₆-tagged XKCM1 (50 nM), or with the purified, bacterially expressed (His)₆-tagged catalytic domain (500 nM) of either KLP10A (amino acids 204–609) or KLP59C (amino acids 158–568).

Drosophila stocks and embryo collection

The following transgenic flies were provided: GFP-histone from R. Saint and W. Sullivan; GFP-tubulin from A. Spradling; GFP-Cid from S. Henikoff; GFP-MeiS332 from T. Orr-Weaver; GFP-Rod from R. Karess; GFP-Polo from C. Sunkel and GFP-DTACC from J. Raff. Fly stocks were maintained and embryos collected as described²⁸.

Embryo microinjection

Microinjections were performed as described²⁸. Rhodamine-labelled tubulin was purchased from Cytoskeleton. Rhodamine-labelled histones were prepared as described²⁹. Control embryos were injected with nonspecific rabbit IgG, GST or buffer. GST, GST-tagged 10A-NT and GST-tagged 59C-NT proteins (see Supplementary Fig. S2c) were purified in PBS and microinjected at a range of concentrations (GST at 15–30 μM and GST-10A-NT, GST-59C-NT and GST-59D-NT at 6–12 μM ; final intracellular concentrations). Rhodamine labelling of GST-KinI NT proteins was performed on glutathione-Sepharose-bound material according to a modified protocol²⁹, and injected at low concentration (GST-10A-NT at 1.8 μM , GST-59C-NT at 1.4 μM ; final intracellular concentrations). Affinity-purified anti-KinI antibodies (see Supplementary Fig. S4) were microinjected at a range of concentrations (anti-KLP10A at 0.16–0.19 mg ml^{-1} , anti-KLP59C at 0.22–0.28 mg ml^{-1} ; final intracellular concentrations). Anti-KLP10A Fabs (from affinity-purified 656 IgG) and anti-KLP59C Fabs (from affinity-purified 696 IgG) were prepared using ImmunoPure Fab Kit (Pierce) and injected at a final concentration of 0.05 and 0.08 mg ml^{-1} , respectively. Cy5 labelling of affinity-purified antibodies was performed using the Zenon Alexa Fluor-647 Labeling Kit (Molecular Probes).

Spinning disk confocal microscopy and FSM

Images and time-lapse movies were acquired using an UltraView Spinning Disk Confocal (PerkinElmer) mounted on a Nikon Eclipse TE 300 inverted microscope with a $\times 60$, 1.4 N.A. objective. Multiple 1- μm z-sections were obtained using a piezo-electric z-axis controller for four-dimensional data set collection. FSM in embryos was performed as previously described¹⁸. Briefly, embryos were injected with dilute concentrations of rhodamine-tubulin (final intracellular concentration of 1–5 $\mu\text{g ml}^{-1}$) and sections of a single z-plane were acquired over time.

Image processing and data analysis

Data sets were saved as stacks of TIFF files. Z-series were saved as maximum intensity projections and time series were saved as AVI movies. Data sets were processed and analysed with MetaMorph (Universal Imaging). Analyses of mitotic phenotypes were restricted to a region approximately one-third of the embryo originating from the site of injection (see Supplementary Fig. S7). Where noted, statistical analysis was performed using a two-tailed, two-sample *t*-test for independent samples after first testing variance equality using an *F* test. FSM images were processed as follows: background was subtracted, regions of interest were passed through a Sharpen high filter, then a 4×4 low-pass filter, and finally, pixel intensities were 'multiplied' 9/6. Point-tracking motion analysis was performed by 'hand' using the Calipers tool. Motion analyses were also performed using the kymograph function. Kinetochore alignment was assessed as follows: the centroid position of all kinetochores on a spindle was averaged and plotted in relation to the centroid of the entire spindle. Note that all data points in these analyses represent the averaged position of all kinetochores on a spindle.

Received 17 September; accepted 28 November 2003; doi:10.1038/nature02256.

Published online 14 December 2003.

- Mitchison, T. J. & Salmon, E. D. Mitosis: a history of division. *Nature Cell Biol.* **3**, E17–E21 (2001).
- Wittmann, T., Hyman, A. & Desai, A. The spindle: a dynamic assembly of microtubules and motors. *Nature Cell Biol.* **3**, E28–E34 (2001).
- Rieder, C. L. & Salmon, E. D. The vertebrate cell kinetochore and its roles during mitosis. *Trends Cell Biol.* **11**, 101–108 (2001).

- Biol.* **8**, 310–318 (1998).
4. Mitchison, T. J., Evans, L., Schulze, E. & Kirschner, M. Sites of microtubule assembly and disassembly in the mitotic spindle. *Cell* **45**, 515–527 (1986).
 5. Gorbisky, G. J., Sammak, P. J. & Borisy, G. G. Chromosomes move poleward in anaphase along stationary microtubules that coordinately disassemble from their kinetochore ends. *J. Cell Biol.* **104**, 9–18 (1987).
 6. Mitchison, T. J. Polewards microtubule flux in the mitotic spindle: evidence from photoactivation of fluorescence. *J. Cell Biol.* **109**, 637–652 (1989).
 7. Hunter, A. W. & Wordeman, L. How motor proteins influence microtubule polymerization dynamics. *J. Cell Sci.* **24**, 4379–4389 (2000).
 8. Wordeman, L. & Mitchison, T. J. Identification and partial characterization of mitotic centromere-associated kinesin, a kinesin-related protein that associates with centromeres during mitosis. *J. Cell Biol.* **128**, 95–104 (1995).
 9. Walczak, C. E., Mitchison, T. J. & Desai, A. XKCM1: a *Xenopus* kinesin-related protein that regulates microtubule dynamics during mitotic spindle assembly. *Cell* **84**, 37–47 (1996).
 10. Desai, A., Verma, S., Mitchison, T. J. & Walczak, C. E. KinI kinesins are microtubule-destabilizing enzymes. *Cell* **96**, 69–78 (1999).
 11. Maney, T., Hunter, A. W., Wagenbach, M. & Wordeman, L. Mitotic centromere-associated kinesin is important for anaphase chromosome segregation. *J. Cell Biol.* **142**, 787–801 (1998).
 12. Kline-Smith, S. L. & Walczak, C. E. The microtubule-destabilizing kinesin XKCM1 regulates microtubule dynamic instability in cells. *Mol. Biol. Cell* **13**, 2718–2731 (2002).
 13. Walczak, C. E., Gan, E. C., Desai, A., Mitchison, T. J. & Kline-Smith, S. L. The microtubule-destabilizing kinesin XKCM1 is required for chromosome positioning during spindle assembly. *Curr. Biol.* **12**, 1885–1889 (2002).
 14. Goldstein, L. S. B. & Gunawardena, S. Flying through the *Drosophila* cytoskeletal genome. *J. Cell Biol.* **150**, F63–F68 (2000).
 15. Sullivan, W. & Theurkauf, W. E. The cytoskeleton and morphogenesis of the early *Drosophila* embryo. *Curr. Opin. Cell Biol.* **7**, 18–22 (1995).
 16. Sullivan, W., Fogarty, P. & Theurkauf, W. E. Mutations affecting the cytoskeletal organization of syncytial *Drosophila* embryos. *Development* **118**, 1245–1254 (1993).
 17. Waterman-Storer, C. M., Desai, A., Bulinski, J. C. & Salmon, E. D. Fluorescent speckle microscopy, a method to visualize the dynamics of protein assemblies in living cells. *Curr. Biol.* **8**, 1227–1230 (1998).
 18. Brust-Mascher, I. & Scholey, J. M. Microtubule flux and sliding in mitotic spindles of *Drosophila* embryos. *Mol. Biol. Cell* **13**, 3967–3975 (2002).
 19. Maddox, P., Desai, A., Oegema, K., Mitchison, T. J. & Salmon, E. D. Poleward microtubule flux is a major component of spindle dynamics and anaphase A in mitotic *Drosophila* embryos. *Curr. Biol.* **12**, 1670–1674 (2002).
 20. Waters, J. C., Mitchison, T. J., Rieder, C. L. & Salmon, E. D. The kinetochore microtubule minus-end disassembly associated with poleward flux produces a force that can do work. *Mol. Biol. Cell* **7**, 1547–1558 (1996).
 21. Sharp, D. J., Rogers, G. C. & Scholey, J. M. Cytoplasmic dynein is required for poleward chromosome movement during mitosis in *Drosophila* embryos. *Nature Cell Biol.* **2**, 922–930 (2000).
 22. Troxell, C. L. *et al.* *klp1⁺* and *klp2⁺*: Two kinesins of the Kar3 subfamily in fission yeast perform different functions in both mitosis and meiosis. *Mol. Biol. Cell* **12**, 3476–3488 (2001).
 23. West, R. R., Malmstrom, T., Troxell, C. L. & McIntosh, J. R. Two related kinesins, *klp5⁺* and *klp6⁺*, foster microtubule disassembly and are required for meiosis in fission yeast. *Mol. Biol. Cell* **12**, 3919–3932 (2001).
 24. West, R. R., Malmstrom, T. & McIntosh, J. R. Kinesins *klp5⁺* and *klp6⁺* are required for normal chromosome movement in mitosis. *J. Cell Sci.* **115**, 931–940 (2002).
 25. Garcia, M. A., Koonrugs, N. & Toda, T. Two kinesin-like KinI family proteins in fission yeast regulate the establishment of metaphase and the onset of anaphase. *Curr. Biol.* **12**, 610–621 (2002).
 26. Rogers, S. L., Rogers, G. C., Sharp, D. J. & Vale, R. D. *Drosophila* EB1 is important for proper assembly, dynamics, and positioning of the mitotic spindle. *J. Cell Biol.* **158**, 873–884 (2002).
 27. Bousbaa, H., Correia, L., Gorbisky, G. J. & Sunkel, C. E. Mitotic phosphoepitopes are expressed in Kc cells, neuroblasts and isolated chromosomes of *Drosophila melanogaster*. *J. Cell Sci.* **110**, 1979–1988 (1997).
 28. Sharp, D. J., Yu, K. R., Sisson, J. C., Sullivan, W. & Scholey, J. M. Antagonistic microtubule-sliding motors position mitotic centrosomes in *Drosophila* early embryos. *Nature Cell Biol.* **1**, 51–54 (1999).
 29. Valdes-Perez, R. E. & Minden, J. S. *Drosophila melanogaster* syncytial nuclear divisions are patterned: time-lapse images, hypothesis and computational evidence. *J. Theor. Biol.* **175**, 525–532 (1995).
 30. McIntosh, J. R., Grishchuk, E. L. & West, R. R. Chromosome-microtubule interactions during mitosis. *Annu. Rev. Cell Dev. Biol.* **18**, 193–219 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank B. Kenigsberg, D. Buster, P. Baas, F. McNally, A. Asenjo and H. Sosa for critically reading the manuscript; I. Brust-Mascher for advice on FSM; and A. Desai and J. Minden for help with rhodamine-histone preparations. We also thank C. Sunkel for providing GFP-Polo kinase flies and protocols for isolating mitotic chromosomes. We are grateful to the following individuals for their gifts: GFP-tubulin flies from A. Spradling; GFP-DTACC flies from J. Raff; GFP-Rod flies from R. Karess; GFP-Cid flies from S. Henikoff; and G. Karpen for the anti-Cid antibody. We thank G. Law, P. Franklin and R. Sleeper (PerkinElmer) for their assistance with the Ultraview confocal microscope. This work was supported by grants from the National Institutes of Health to D.J.S., C.E.W., R.D.V. and J.M.S. C.E.W. is a Scholar of the Leukemia and Lymphoma Society.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.S. (dsharp@aecom.yu.edu).

The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmoreponsive genes

Eulàlia de Nadal¹, Meritxell Zapater¹, Paula M. Alepuz^{1*}, Lauro Sumoy², Glòria Mas¹ & Francesc Posas¹

¹Cell Signaling Unit, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra (UPF), and ²Centre de Regulació Genòmica (CRG), E-08003 Barcelona, Spain

* Present address: Departamento de Bioquímica y Biología Molecular, Universidad de Valencia, E-46100 Valencia, Spain

Regulation of gene expression by mitogen-activated protein kinases (MAPKs) is essential for proper cell adaptation to extracellular stimuli. Exposure of yeast cells to high osmolarity results in rapid activation of the MAPK Hog1, which coordinates the transcriptional programme required for cell survival on osmolestress¹. The mechanisms by which Hog1 and MAPKs in general regulate gene expression are not completely understood, although Hog1 can modify some transcription factors². Here we propose that Hog1 induces gene expression by a mechanism that involves recruiting a specific histone deacetylase complex to the promoters of genes regulated by osmolestress. Cells lacking the Rpd3–Sin3 histone deacetylase complex are sensitive to high osmolarity and show compromised expression of osmolestress genes. Hog1 interacts physically with Rpd3 *in vivo* and *in vitro* and, on stress, targets the deacetylase to specific osmolestress-responsive genes. Binding of the Rpd3–Sin3 complex to specific promoters leads to histone deacetylation, entry of RNA polymerase II and induction of gene expression. Together, our data indicate that targeting of the Rpd3 histone deacetylase to osmolestress-responsive promoters by the MAPK Hog1 is required to induce gene expression on stress.

In response to high osmolarity, Hog1 coordinates the transcriptional programme required for cell adaptation, and thus *hog1Δ* cells are osmosensitive³. Hog1 tightly binds chromatin in response to osmolestress and elicits gene transcription by mechanisms other than the simple modification of activators^{4,5}. To test whether the Hog1 MAPK can affect chromatin organization, we screened for osmosensitive mutants on the basis that mutations in chromatin-modifying activities that affect osmolestress gene expression should result in cells that are osmosensitive.

Deleting genes encoding the Rpd3 histone deacetylase and its interacting protein Sin3 rendered cells osmosensitive (Fig. 1a). Notably, strains containing the double mutation *hog1Δ rpd3Δ* or *hog1Δ sin3Δ* were as osmosensitive as the single *hog1Δ* strain (data not shown). Rpd3 is a member of a family of five related histone deacetylases in yeast that also comprises Hda1, Hos1, Hos2 and Hos3. Deleting the genes encoding these related deacetylases did not affect cellular osmosensitivity (Fig. 1a and data not shown).

As histone deacetylases are involved in regulating gene expression, we tested whether the expression of Hog1-dependent genes was affected by deleting *RPD3*. Time course experiments showed that the expression of various osmolestress-inducible genes, such as *HSP12*, *STL1*, *CTT1*, *ALD3*, *ARO9*, *ENA1* and *GRE2*, was reduced in the *rpd3Δ* strain on stress (Fig. 1b and not shown). Thus, Hog1-mediated gene expression is impaired by deleting *RPD3*.

To assess the relevance of Rpd3 in Hog1-mediated gene expression, we used microarray analysis to compare gene induction in wild type and *rpd3Δ* strains on osmolestress. A high proportion of the genes induced on stress was affected by deleting the deacetylase. Of the 222 genes induced more than twofold in response to stress, more than 90% showed a significant reduction in expression after a

short period of stress (5 min, 0.4 M NaCl) and about 80% of the genes were affected after longer exposures to stress (20 min, 0.4 M NaCl; Supplementary Fig. S1). Thus, Rpd3 histone deacetylase is important for eliciting gene expression in response to osmotic stress.

The presence of Hog1 in osmoresponsive promoters is important for inducing gene expression because it facilitates the recruitment of components of the transcription initiation complex⁴⁻⁶. In fact, when tethered to a promoter as a LexA fusion protein, Hog1 can recruit the RNA polymerase II (Pol II) complex and induce transcription from a LexA-*lacZ* reporter system on osmostress⁵. We therefore tested whether activation by LexA-Hog1 would be affected by deleting the deacetylase complex. Gene induction by LexA-Hog1 on osmostress was completely impaired in cells deficient in *RPD3* or *SIN3*, but not in cells deficient in other deacetylases (Fig. 2a).

To determine whether the function of Rpd3 in osmostress requires its histone deacetylase activity, we created a point mutant allele (in which His 150 and His 151 were changed to alanine; denoted 150:151) that has been shown to abolish Rpd3 deacetylase

activity⁷. The catalytically impaired Rpd3 mutant did not restore expression of the LexA-Hog1 in the *rp3Δ* strain (Fig. 2a) and was unable to complement *rp3Δ* osmosensitivity (Fig. 1a), thereby indicating that proper Hog1-mediated transcription responses require Rpd3 catalytic activity.

Rpd3 mainly deacetylates acetyl groups in histones H3 and H4 (ref. 8). To test whether histone acetylation is a key step for gene induction, we analysed gene expression induced by LexA-Hog1 in strains containing modified histone H4 acetylation sites⁹. Replacing the lysines in histone H4 with arginine (hereafter denoted K-R) has been shown to mimic the deacetylated state of histone H4, whereas replacing them with glutamine (K-Q) resembles histone H4 acetylation⁹. Expression by LexA-Hog1 was increased in the H4(K-R) mutant strain and, more relevantly, it was completely abolished in the histone H4(K-Q) mutant, as well as in the *rp3Δ* strain (Fig. 2b).

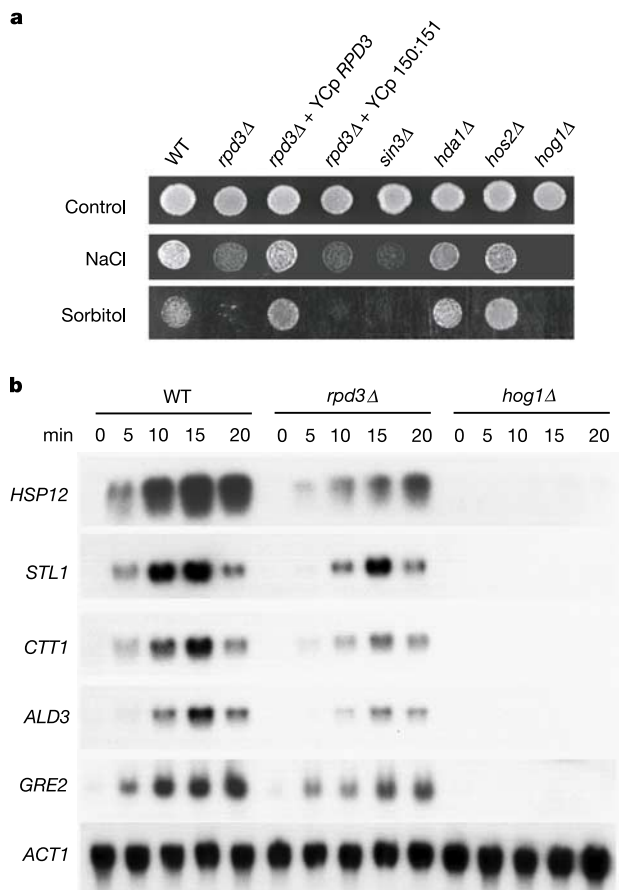


Figure 1 Mutations in *RPD3* and *SIN3* render cells osmosensitive with impaired expression of osmostress genes. **a**, Effects on the sensitivity to osmolarity of several histone deacetylase mutants. The wild-type (TM101) strain and its derivatives *rp3Δ*, *sin3Δ*, *hda1Δ*, *hos2Δ* and *hog1Δ* were spotted on YPD plates with or without 1.2 M NaCl or 2.4 M sorbitol. *rp3Δ* mutant cells were transformed with centromeric plasmids expressing wild-type *RPD3* (YCp *RPD3*) or a catalytically inactive allele (YCp 150:151). Growth was scored after 4 d. **b**, Impaired expression of osmostress genes in a *rp3Δ* strain. Wild-type, *rp3Δ* and *hog1Δ* cells were grown to mid-log phase in rich medium and then subjected to osmotic shock (0.4 M NaCl) for the indicated times. Total RNA was assayed by northern blot for *HSP12*, *STL1*, *CTT1*, *ALD3*, *GRE2* and *ACT1* (as a loading control).

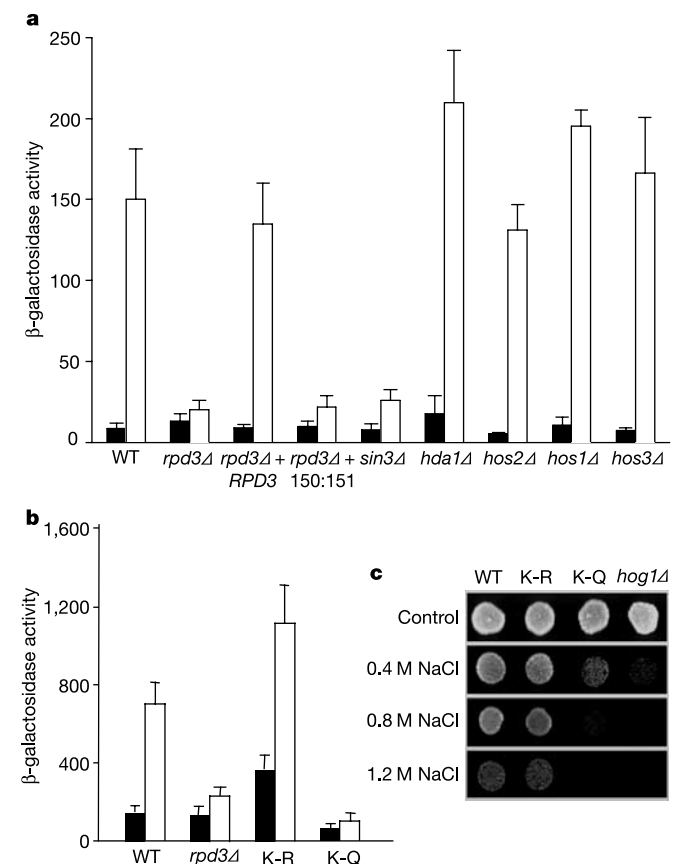


Figure 2 Osmostress gene expression mediated by LexA-Hog1 is affected by deleting *RPD3* and *SIN3* or by modifying histone H4 lysines. **a**, Transcriptional activation by LexA-Hog1 is abolished in *RPD3* and *SIN3*. L40 cells (a LexA-*lacZ* reporter strain) with *rp3Δ*, *sin3Δ*, *hda1Δ*, *hos1Δ*, *hos2Δ* and *hos3Δ* mutations were transformed with pBTM116-Hog1. *rp3Δ* mutant cells were transformed with centromeric plasmids expressing wild-type *RPD3* (+*RPD3*) or a catalytically inactive allele (+150:151). β -Galactosidase activity was assayed in cells that were grown to mid-log phase and either untreated (filled bars) or subjected (open bars) to hyperosmotic stress (0.4 M NaCl for 30 min). β -Galactosidase values are given in nmol min⁻¹ mg⁻¹. **b**, Mutations in histone H4 lysines affect expression by the LexA-Hog1 in response to stress. Wild-type cells or strains with modified histone H4 tails were transformed with pBTM116-Hog1 and pH18-8 (LexA-*lacZ*) plasmids. In the K-R strain, lysines 5, 8, 12 and 16 are changed to arginine⁸; in the K-Q strain, lysines 5, 8, 12 and 16 are changed to glutamine⁹. **c**, Cells in which the amino-terminal lysine of histone H4 is changed to glutamine are osmosensitive. The strains described in **b** were spotted on YPD plates with or without 1.2 M NaCl. Growth was scored after 4 d.

Expression induced by LexA-Hog1 was higher in a H4(K-R) mutant strain containing the *RPD3* deletion than in the *rp3Δ* strain, but lower than in the parental strain, suggesting that the effects of Rpd3 are not restricted to histone H4, as previously described⁸. Expression by a LexA-Gcn4 activator was unaffected in the H4(K-R) or H4(K-Q) strains (data not shown). Thus, the degree of histone acetylation is important for Hog1-mediated gene expression. Consistent with this observation, we found that the histone H4(K-Q) mutant strain is osmosensitive (Fig. 2c).

Binding of Hog1 to osmoresponsive promoters is a key step in Hog1-mediated gene expression. We therefore examined whether Hog1 physically associates with the deacetylase to localize its function to specific promoters. In *in vivo* pull-down experiments, Hog1 coprecipitated Rpd3 and Sin3, but not the related Hda1 deacetylase (Fig. 3a, c). Moreover, in *in vitro* experiments using purified proteins, Rpd3 directly interacted with Hog1 (Fig. 3b). Therefore, Hog1 physically interacts with the Rpd3-Sin3 deacetylase complex.

Because Hog1 interacts with Rpd3 and is bound to promoters only in response to osmotic stress, we next examined whether Rpd3 is

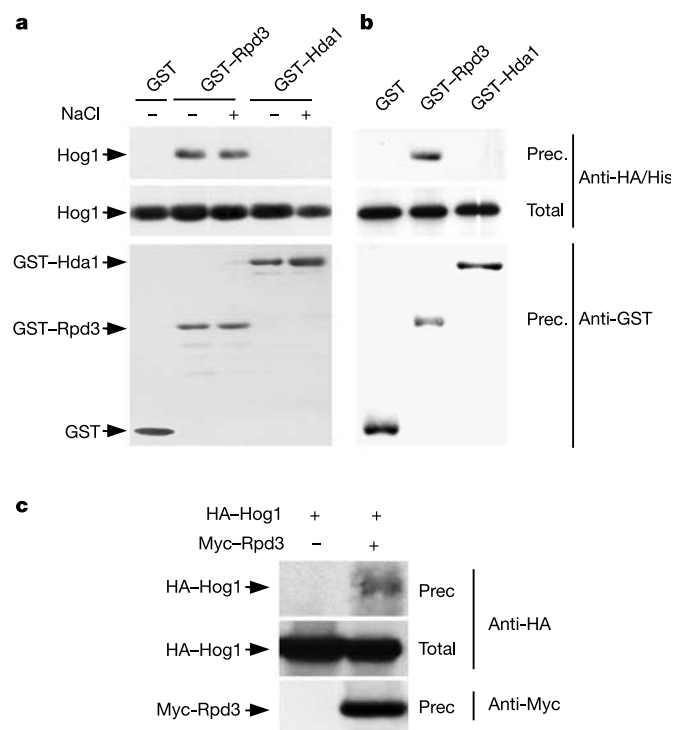


Figure 3 Hog1 physically interacts with Rpd3. **a**, Hog1 coprecipitates with Rpd3 *in vivo*. A strain expressing HA-Hog1 from the wild-type locus was transformed with a plasmid expressing GST, GST-Rpd3 or GST-Hda1 under the *P_{TEF1}* promoter. After growth, cells were either untreated (–) or subjected (+) to osmotic stress (10 min, 0.4 M NaCl). GST proteins were pulled down by glutathione-Sepharose, and GST-containing proteins (top) and HA-Hog1 (bottom) were detected by western blotting using antibodies against GST and HA. Total represents <20% of total input protein (middle). **b**, Hog1 associates directly with Rpd3 *in vitro*. GST, GST-Rpd3, GST-Hda1 and His-Hog1 were expressed and purified from *E. coli*. Hog1 was incubated with GST fusion proteins and was detected by immunoblotting using His antibody (top). Total represents <20% of total input protein (middle). **c**, Strains expressing HA-Hog1 (+) and Myc-Rpd3 (+) or not (–) from the chromosomal locus were grown and subjected to osmotic stress as in **a**. Myc-tagged proteins were immunoprecipitated by monoclonal antibodies against Myc, and Myc-Rpd3 protein (top) and HA-Hog1 (bottom) were detected by western blotting using antibodies against Myc and HA, respectively. Total represents <50% of total input protein (middle).

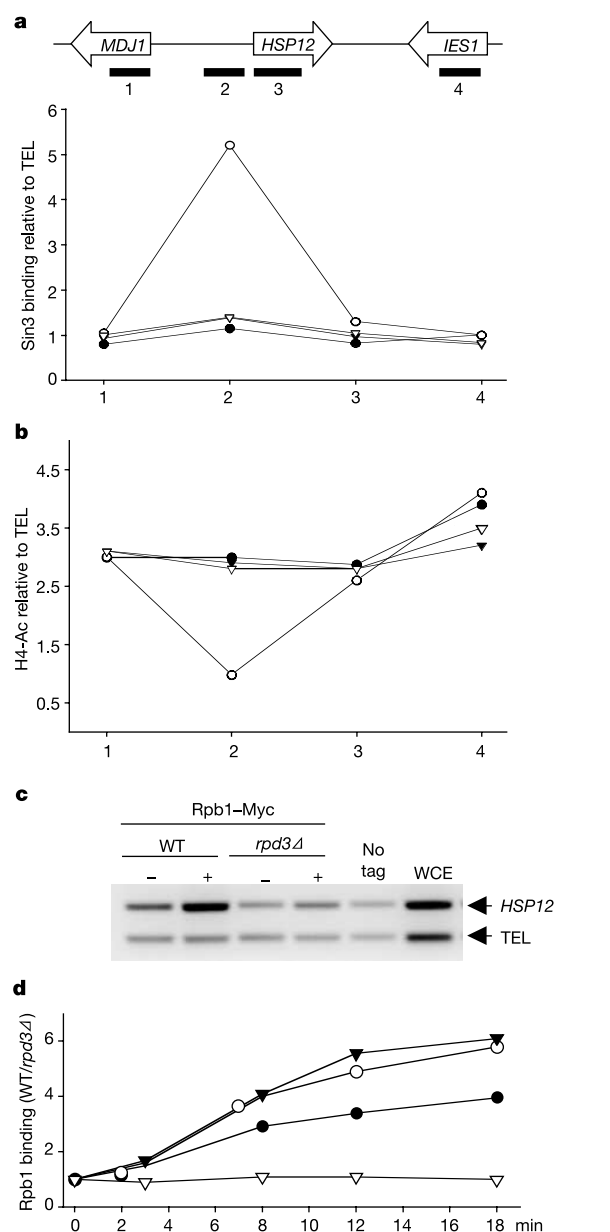


Figure 4 The Rpd3-Sin3 complex directly binds and deacetylates the *HSP12* promoter to facilitate Pol II binding. **a**, Binding of Sin3 to the chromosomal region containing *HSP12* depends on stress and Hog1. Wild-type (circles) or *hog1Δ* (triangles) strains expressing genomic Sin3-HA were grown, and samples for ChIP assay were either untreated (filled symbols) or subjected (open symbols) to osmotic stress (10 min, 0.4 M NaCl). The relative intensity of PCR fragments normalized to the telomere-specific band (internal control) is shown. The location of the primers used for PCR is shown at the top. Data from one representative experiment out of three are shown. **b**, Histone H4 acetylation at the *HSP12* promoter is decreased on osmotic stress. ChIP assays were done in wild-type (circles) or *hog1Δ* (triangles) cells with an antibody against acetylated histone H4. Cells were subjected to osmotic stress as in **a**. **c**, Recruitment of RNA Pol II to *HSP12* occurs in response to osmotic stress and depends on Rpd3 deacetylase. Cross-linked cell extracts from non-stressed (–) or osmotically stressed (+) wild-type and *rp3Δ* strains expressing chromosomal Rpb1-Myc were immunoprecipitated by anti-Myc antibodies, and assayed for the presence of the *HSP12* promoter or TEL regions. As controls, DNA was amplified from extracts without tagged Rpb1 ('no tag'), and whole-cell extract (WCE) was assayed before immunoprecipitation. **d**, Time course of the recruitment of RNA Pol II to Rpd3-dependent promoters. Experiments were carried out essentially as in **c** at the indicated times after NaCl addition (0.4 M NaCl). Binding to *STL1* (filled triangles), *ALD3* (open circles), *CTT1* (filled circles) and the control *GAL1* (open triangles) was assayed, and results are shown as the ratio of RNA Pol II binding in wild type to binding in the *rp3Δ* strain.

recruited to osmostress promoters by Hog1 in response to stress. Chromatin immunoprecipitation (ChIP) analysis showed that on osmostress Sin3 and Rpd3 associate specifically (an above fivefold enrichment) with the promoter region of the *HSP12*, whereas no increase in binding was observed in the surrounding regions. Furthermore, no binding was observed on stress in a *hog1Δ* strain (Fig. 4a and data not shown). Thus, binding of the Rpd3–Sin3 deacetylase complex occurs only at the *HSP12* promoter region and depends on osmostress and the presence of Hog1.

We tested whether acetylation of the *HSP12* promoter is affected by recruitment of the Rpd3–Sin3 deacetylase complex on osmostress, by using ChIP analysis with antibodies specific for acetylated histone H4. A decrease in histone acetylation was observed only at the promoter region of *HSP12* and only on stress in a wild-type strain, whereas no differences in acetylation were observed in the *hog1Δ* strain (Fig. 4b). Thus, on osmostress specific recruitment of the Rpd3–Sin3 deacetylase complex to the promoter region causes a decrease in the acetylation of histone H4.

We also tested whether the presence of the deacetylase complex and the decrease in promoter acetylation are associated with gene activation by the RNA Pol II complex at the *HSP12* promoter. Analysis of Rpb1 binding by ChIP analysis showed that Rpb1 is associated with the *HSP12* promoter only in response to osmostress and that Rpb1 binding is strongly reduced in *hog1Δ* and *rp3Δ* strains (Fig. 4c and not shown), consistent with the fact that *HSP12* gene expression is reduced in *hog1Δ* and *rp3Δ* strains on stress. As expected from the expression pattern (Fig. 1b), recruitment of Rpd3 to the *STL1*, *CTT1*, *ALD3* and *GRE2* promoters also occurred on osmostress and in a *HOG1*-dependent manner (Supplementary Fig. S2b and data not shown). But Rpd3 was not recruited to other genes, such as *YLR164w*, that were induced in response to stress but whose expression was not affected by deleting *RPD3* or *HOG1* (data not shown). As expected, recruitment of RNA Pol II to those genes whose expression was *RPD3*-dependent was markedly reduced in the *rp3Δ* strain (Fig. 4d). Together, these data indicate that Rpd3 has a positive role in gene expression.

Histone deacetylation has been traditionally associated with the repression of gene expression. Although on a global scale this seems to be true¹⁰, there are many examples of particular genes in which a decrease in histone acetylation is associated with transcription induction^{11–13}. We thus propose the following working model for Rpd3-mediated gene expression in osmostress. In response to stress, Hog1 is recruited to osmostress promoters by specific transcription factors⁶. Hog1 binding facilitates direct recruitment of the Rpd3 deacetylase complex to the promoters, leading to histone deacetylation, entry of RNA Pol II and the induction of gene expression. Although the deacetylation of histones seems to be essential for gene induction by Rpd3, we cannot exclude the possibility that other non-histone substrates of Rpd3 may contribute to gene induction.

Several genes induced by osmostress also respond to other stresses. For example, *CTT1* and *HSP12* are also induced by heat stress, but their induction by heat is independent of Hog1. Because Rpd3 is essential for inducing these genes on osmostress, we considered that Rpd3 may be required to induce them on heat stress. We therefore analysed the role of Rpd3 in *CTT1* and *HSP12* gene expression on heat shock. Expression of *CTT1* and *HSP12* was induced rapidly by heat stress in a *HOG1*-independent manner. Notably, induction of *HSP12* and *CTT1* was also impaired on heat shock in a *rp3Δ* strain (Supplementary Fig. S2a). Thus, Rpd3 is required to induce gene expression on osmostress and heat stress, whereas Hog1 is only involved in osmostress-induced gene expression.

As Rpd3 is present at the promoters that require Rpd3 for proper osmostress gene expression, we reasoned that Rpd3 might be also recruited to the *CTT1* and *HSP12* promoters on heat stress. Indeed, ChIP analysis showed that the deacetylase complex binds to the

CTT1 promoter in response to heat stress, as well as osmostress, in a wild-type strain (Supplementary Fig. S2b). In agreement with the expression data, recruitment of the deacetylase complex to *CTT1* was abolished in a *hog1Δ* strain in response to osmostress, but still occurred in a *hog1Δ* strain in response to heat stress (Supplementary Fig. S2b). As expected from the requirement of Rpd3 for gene expression on heat stress, deleting *RPD3* or *SIN3* but not *HOG1* resulted in cells that were temperature sensitive (Supplementary Fig. S2c).

Data from the microarray analysis showed that on stress a few genes were induced independently of *HOG1* but were affected by the deletion of *RPD3*. We therefore predicted that recruitment of Rpd3 to those promoters would occur on stress and independently of the presence of Hog1. Indeed, ChIP analysis showed that Rpd3 bound to *YKR033c* or *RPB11* in response to osmostress and independently of *HOG1* (data not shown). Thus, Rpd3 is essential for inducing gene expression in response to stress and is actively recruited by different signalling molecules to responsive promoters depending on the stimuli, Hog1 being the chief targeting molecule of Rpd3 in response to osmostress.

A positive role for deacetylases in gene expression has been also proposed for Hos2, a deacetylase of the same family as Rpd3 that was initially associated with transcription repression⁷. Through work using open reading frames, Hos2 has been also shown to be essential for inducing the expression of *INO1* and *GAL13*. We therefore propose that Rpd3 functions positively in stress gene expression and that on osmostress is an important part of the signalling pathway of the MAPK Hog1. The role of Rpd3 deacetylation in osmostress promoters not only might be restricted to altering chromatin structure but also might provide unique binding surfaces or recognition motifs for the recruitment of activators, as has been proposed for acetylation and deacetylation in gene expression^{8,14,15}. □

Methods

Microarray analysis

Wild-type and *rp3Δ* strains were untreated or treated by osmostress (5–20 min, 0.4 M NaCl) and RNA was extracted with hot phenol. We carried out two-colour microarray hybridizations using glass microarrays representing the whole genome of *Saccharomyces cerevisiae* prepared by the University Health Network (Toronto, Canada). Each comparison was done in triplicate in a total of six hybridization experiments. An intensity-dependent normalization was applied by the lowest method with $f = 20\%$ for all, and replicate spot mean values of normalized $\log_2$ ratios were calculated for each hybridization. The median value of three hybridization experiments was used as the final measure of difference in gene expression between treated and untreated samples for each genotype. Data from these experiments have been submitted to the ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress/>) in compliance with MIAME (minimal information about a microarray experiment). Additional information is given in the Supplementary Information.

ChIP assay

We did ChIP polymerase chain reaction (PCR) assays as described⁶. The antibody against acetylated histone H4 was from Upstate Biotech. In all ChIP experiments, yeast cultures were grown to early log phase (an absorbance of 0.6–1.0 at 600 nm) before cells were exposed to osmotic stress. We used oligonucleotides to amplify chromosomal region 1 (–1,292 to –1,040), 2 (–380 to –128), 3 (–137 to +112) and 4 (+1,027 to +1,324) of *HSP12*. TEL corresponds to a 177-bp region located 500 bp before the chromosome VI right end.

Yeast strains and DNA constructs

A full list of strains and plasmids is given in the Supplementary Information. Yeast strains were derived from W303 or S288C. We verified deletion strains by PCR, and generated plasmids and carried out mutagenesis by standards methods.

Recombinant protein expression

His-tagged Hog1 was expressed in *Escherichia coli* BL21(DE3) cells, purified by TALON metal affinity resin (Clontech) and eluted using imidazole buffer. Glutathione S-transferase (GST) fusion proteins encoding Hda1 and Rpd3 were constructed in pGEX vectors (Pharmacia), expressed in *E. coli* DH5, and purified by glutathione–Sepharose beads as described¹⁶.

Received 27 October; accepted 26 November 2003; doi:10.1038/nature02258.

1. Hohmann, S. Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol. Mol. Biol. Rev.* **66**, 300–372 (2002).

2. de Nadal, E., Alepuz, P. M. & Posas, F. Dealing with osmostress through MAP kinase activation. *EMBO Rep.* **3**, 735–740 (2002).
3. Brewster, J. L., de Valoir, T., Dwyer, N. D., Winter, E. & Gustin, M. C. An osmosensing signal transduction pathway in yeast. *Science* **259**, 1760–1763 (1993).
4. Proft, M. & Struhl, K. Hog1 kinase converts the Sko1–Cyc8–Tup1 repressor complex into an activator that recruits SAGA and SWI/SNF in response to osmotic stress. *Mol. Cell* **9**, 1307–1317 (2002).
5. Alepuz, P. M., de Nadal, E., Zapater, M., Ammerer, G. & Posas, F. Osmostress induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II. *EMBO J.* **22**, 2433–2442 (2003).
6. Alepuz, P. M., Jovanovic, A., Reiser, V. & Ammerer, G. Stress-induced MAP kinase Hog1 is part of transcription activation complexes. *Mol. Cell* **7**, 767–777 (2001).
7. Kadosh, D. & Struhl, K. Histone deacetylase activity of Rpd3 is important for transcriptional repression *in vivo*. *Genes Dev.* **12**, 797–805 (1998).
8. Kurdistani, S. K. & Grunstein, M. Histone acetylation and deacetylation in yeast. *Nature Rev. Mol. Cell. Biol.* **4**, 276–284 (2003).
9. Durrin, L. K., Mann, R. K., Kaye, P. S. & Grunstein, M. Yeast histone H4 N-terminal sequence is required for promoter activation *in vivo*. *Cell* **65**, 1023–1031 (1991).
10. Robyr, D. *et al.* Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* **109**, 437–446 (2002).
11. Deckert, J. & Struhl, K. Histone acetylation at promoters is differentially affected by specific activators and repressors. *Mol. Cell. Biol.* **21**, 2726–2735 (2001).
12. Bernstein, B. E., Tong, J. K. & Schreiber, S. L. Genomewide studies of histone deacetylase function in yeast. *Proc. Natl Acad. Sci. USA* **97**, 13708–13713 (2000).
13. Wang, A., Kurdistani, S. K. & Grunstein, M. Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science* **298**, 1412–1414 (2002).
14. Berger, S. L. Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* **12**, 142–148 (2002).
15. Fischle, W., Wang, Y. & Allis, D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
16. Bilsland-Marchesan, E., Arino, J., Saito, H., Sunnerhagen, P. & Posas, F. Rck2 kinase is a substrate for the osmotic stress-activated mitogen-activated protein kinase Hog1. *Mol. Cell. Biol.* **20**, 3887–3895 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Carmona and D. Otero for technical assistance; J. Lozano for statistical treatment of the microarray data; and M. Beato, F. X. Real and G. Ammerer for advice. M.Z. is the recipient of a Ramón Areces Ph.D. fellowship and G.M. is the recipient of a Formación de Personal Universitario fellowship (MECD, Spanish government). E.d.N. and P.M.A. are supported by the Programa Ramón y Cajal. This work was supported by grants from the Ministerio de Ciencia y Tecnología, 'Distinció de la Generalitat de Catalunya per a la Promoció de la Recerca Universitaria, Joves Investigadors' DURSI (Generalitat de Catalunya), and the European Molecular Biology Organization YIP program to F.P.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.P. (francesc.posas@upf.edu).

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Netherlands/
Italy/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

Shifting the balance

No one would argue that scientific jobs are dispersed unequally around the world. Western, developed countries, especially the United States and Britain, have a disproportionately larger share of funds, both in the public and private sector. As a result, they can attract scientists from parts of the world that have fewer resources or opportunities. There are long- and short-term views of the ramifications of this brain drain from poorer nations — and short- and long-term ways to deal with it.

The older, short-term view is that this is a crisis that requires outside intervention — perhaps investment of vast sums of money by monolithic entities such as the World Bank to allow the nations to build up their scientific infrastructure. A new view, explored in a collection of articles at SciDev.net, an online resource for science in developing countries, says that the brain drain is really 'brain circulation', and that a combination of market forces, good will and sound policy in the poorer nations can create scientific jobs there.

Market forces have already led Indian information-technology experts back from Silicon Valley to Bangalore. Good will — especially when bolstered by philanthropic organizations such as the Bill and Melinda Gates Foundation — is leading Western science to pursue clinical research of AIDS and other infectious diseases in developing countries (see *Nature* 426, 736–737; 2003). And sound policy — in the form of repatriation grants, and a peer-review system free of corruption — is already being used in European countries that have lost scientists to the United States, and could easily be adopted by poorer nations.

It's likely that well-intentioned bodies will try both short- and long-term approaches. But the scientists will flow, as always, to where they have the most opportunities and fewest barriers to entry.

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Geography's brave
new world p376

CAREER VIEW

Scientists & Societies

Eurodoc

Graduate Journal

A tough challenge

Movers

Otmar Wiestler p378

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS



Mapping opportunities

Scientists who can combine geographic information systems with satellite data are in demand in a variety of disciplines. Virginia Gewin gets her bearings.

Forest fires ravaging southern California, foot-and-mouth disease devastating the British livestock industry, the recent outbreak of severe acute respiratory syndrome (SARS) — all of these disasters have at least one thing in common: the role played by geospatial analysts, mining satellite images for information to help authorities make crucial decisions. By combining layers of spatially referenced data called geographic information systems (GIS) with remotely sensed aerial or satellite images, these high-tech geographers have turned computer mapping into a powerful decision-making tool.

Natural-resource managers aren't the only ones to take notice. From military planning to real estate, geospatial technologies have changed the face of geography and broadened job prospects across public and private sectors.

Earlier this year, the US Department of Labor identified geotechnology as one of the three most important emerging and evolving fields, along with nanotechnology and biotechnology. Job opportunities are growing and diversifying as geospatial technologies prove their value in ever more areas.

The demand for geospatial skills is growing worldwide, but the job prospects reflect a country's geography, mapping history and even political agenda. In the United States, the focus on homeland security has been one of many factors driving the job market. Another is its vast, unmapped landscape. While European countries are integrating GIS into government decision-making, their well-charted lands give them little need for expensive satellite imagery.

AN EXPANDING MARKET

All indications are that the US\$5-billion worldwide geospatial market will grow to \$30 billion by 2005 — a dramatic increase that is sure to create new jobs, according to Emily DeRocco, assistant secretary at the US Department of Labor's employment and training division. NASA says that 26% of its most highly trained geotech staff are due to retire in the next decade, and the National Imagery and Mapping Agency is expected to need 7,000 people trained in GIS in the next three years.

Of the 140,000 organizations globally that use GIS,

most are government agencies — local, national and international. A ten-year industry forecast put together last year by the American Society for Photogrammetry & Remote Sensing (ASPRS) identified environmental, civil government, defence and security, and transportation as the most active market segments.

Business at the Earth-imagery provider Space Imaging, of Thornton, Colorado, increased by 70% last year, says Gene Colabatistto, executive vice-president of the company's consulting service. To keep up momentum, the company plans to hire more recruits with a combination of technical and business skills. Colabatistto cites the increased adoption of GIS technologies by governments as a reason for the rise. He adds that the US military, the first industry to adopt GIS and remote sensing on a large scale, has spent more than \$1 billion on commercial remote sensing and GIS in the past two years.

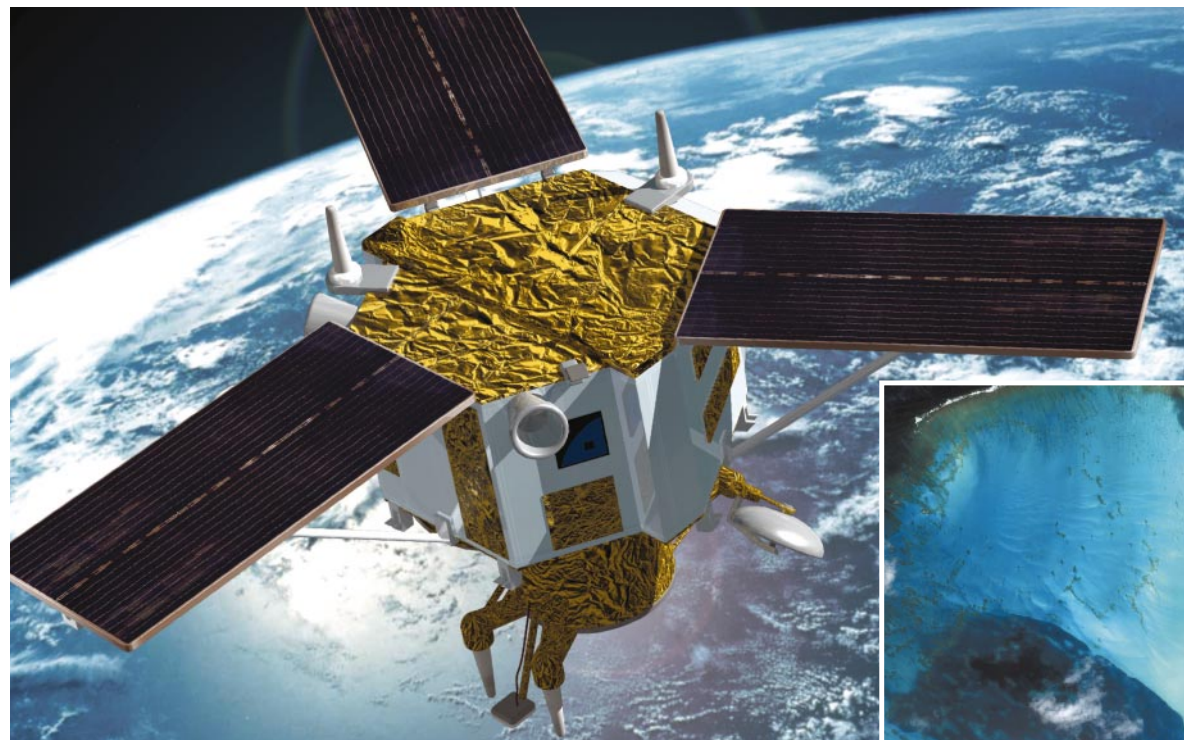
LOOKING DOWN IS LOOKING UP

The private sector hasn't traditionally offered many jobs for geographers, but location-based services and mapping — or 'geographic management systems' — are changing the field. "The business of looking down is looking up," says Thomas Lillesand, director of the University of Wisconsin's Environmental Remote Sensing Center in Madison, Wisconsin.

Imagery providers such as Digital Globe of Longmont, Colorado, also need more GIS-trained workers as markets continue to emerge. Spokesman Chuck Herring says that the company has identified 54 markets in which spatial data are starting to play a role.

The Environmental Systems Research Institute (ESRI), in Redlands, California, sets the industry standards for geospatial software. Most of its 2,500 employees have undergraduate training in geography or information technology, although PhDs are sought after to fill the software-development positions. Many private companies, including the ESRI and Space Imaging, offer valuable work experience to both graduate and final-year undergraduate students.

Graduates in natural-resource management note that GIS and remote-sensing skills are becoming as important as fieldwork. GIS platforms, which manipulate all forms of image data, are transforming



A fresh view: high-resolution imaging satellites such as IKONOS are providing researchers with unparalleled insight and information about the world (inset).

disciplines such as ecology, marine biology and forestry.

"Science has discovered geography," says Doug Richardson, executive director of the Association of American Geographers (AAG). Many of the National Science Foundation's multidisciplinary research programmes now include a geospatial component.

SKILLED LABOUR

Some universities are offering two-year non-thesis master's programmes in geospatial technologies, including communication and business courses — perfect for professionals who want to build on existing skills or move into a new field. The non-profit Sloan Foundation has funded several geospatially related professional master's programmes. In addition, numerous short courses are available to bring professionals up to speed. Indeed, the ESRI alone trains over 200,000 people a year. AAG and ASPRS conferences also offer training sessions.

Although technical skills are important, Richardson stresses that employees need a deep understanding of underlying geographic concepts. "It's a mistake to think that these technologies require only technician-oriented functions," he says.

Throughout the European Union (EU), the many top-quality graduate geography programmes remain the primary training grounds. Recently, a few pan-European projects have also emerged, including a new international institute designed to train future geographers. Building on a collaboration between the European Space

Agency and the US National Science Foundation, the Vespucci Initiative in 2002 began three-week summer workshops training students from around the world in spatial data infrastructure, spatial analysis and geodemographics. The EU even promotes distance learning: UNIGIS, a network of European universities, prides itself on being the only virtual, global, multilingual GIS programme in the world.

Although GIS is increasingly incorporated into UK government practices, there is little demand for remote-sensing expertise in this small and heavily mapped country. Mark Linehan, director of the London-based Association for Geographic Information, says that although the public-sector market is growing, it remains a struggle to find jobs for MScs at the appropriate pay scale and qualification level.

The European Commission (EC) is laying the groundwork to ease data-sharing across countries in anticipation of wider adoption of GIS among the member-state governments and to cut the costs of data gathering. That process alone will require at least a couple of thousand people trained in GIS, and many more proposals are expected.

Indeed, the EC and the European Space Agency have joined to propose a Global Monitoring for Environment and Security initiative, to provide permanent access to information on environmental management, risk surveillance and civil security. Given the scope of the mandate, this is likely to need people who understand how to interpret, integrate and manage satellite information — those who also have a background in natural-resource issues will be in highest demand.

Considering the role that GIS played in staving the spread of foot-and-mouth disease, such a system will not only increase the prevalence of geospatial skills in Europe, it will better connect data with Europe's resource managers.

Virginia Gewin is a freelance science writer in Corvallis, Oregon.

Web links

- Environmental Systems Research Institute
- ♦ www.esri.com
- Association of American Geographers
- ♦ www.aag.org
- American Society for Photogrammetry & Remote Sensing
- ♦ www.asprs.org
- The Vespucci Initiative
- ♦ www.vespucci.org
- Global Monitoring for Environment and Security
- ♦ intelligence.jrc.cec.eu.int/space/baveno/baveno.html
- Association for Geographic Information
- ♦ www.agi.org.uk
- UNIGIS
- ♦ www.unigis.at/net
- US Department of Labor's Career Voyages
- ♦ www.careervoyages.gov
- Sloan Foundation professional master's programmes
- ♦ www.sciencemasters.com



Doug Richardson sees a combination of technological skills and an understanding of geographical concepts as important.

GRADUATE JOURNAL

A tough challenge

Despite physics being the subject I found most difficult at school, or maybe because it was, I decided to pursue it further. I spent four years at the University of Cambridge, UK, and then moved to London to start my PhD. For the past 18 months I have been based at Fermilab, a particle accelerator in the United States, near Chicago. Fermilab currently has the highest-energy accelerator in the world and is an exciting place to work.

At Fermilab, we ask the most fundamental questions of the Universe. Where do we come from? How was the Universe created? How will it end? These are some of the most challenging questions humanity can ask, and are for me the most inspiring ones.

But I also want a job that provides a tangible benefit to society and that makes a difference to people, if only in a small way. Although I feel enormously lucky to do what I do, I also recognize that my research is very esoteric. I would enjoy a career that allows me to share some of my passion for science with others, for example to walk into a classroom and excite and enthuse children about what Fermilab does. My biggest decision after finishing my PhD in the next year is whether to continue in my field of research, or to choose something that touches society a little more directly. ■

Amber Jenkins is a graduate student in particle physics at Imperial College London.

SCIENTISTS & SOCIETIES

Giving young European students a voice

Eurodoc is a European federation of national organizations that represent young researchers in their respective countries. It aims to give a voice to PhD candidates, postdoctoral fellows and other junior researchers in European policy-making processes and to improve their working conditions. To achieve these goals, Eurodoc set up several workgroups that operate open mailing lists and a discussion forum.

So far, postdoc and graduate-student associations from 13 countries are members of Eurodoc, with more expressing interest in joining. Eurodoc is also encouraging young scientists in other countries to establish their own national organizations.

Both board and working-group members regularly attend

conferences and deliver talks, poster presentations or position papers to policy-makers in their field of research in Europe. In addition, Eurodoc organizes an annual conference, gathering the national delegates with guest speakers from policy, academia and industry, for discussions at plenary sessions and thematic workshops.

The next annual conference will be held in Athens on 18–21 March. The major topics to be addressed are the profession of the researcher in the European Research Area, with a keynote speech by Achilleas Mitsos, director-general of the European Commission's research directorate, and the doctoral level as the third tier of higher education, with Eric Froment, president of the European University Association.

As well as this event, Eurodoc is currently co-

organizing a conference on early-stage researcher's mobility, to take place in Lisbon on 25–27 February. The results are expected to significantly raise awareness of this issue and help to remove the national and European obstacles to mobility within and beyond the European Research Area. Eurodoc will also be a member of the steering committee of the research project called 'Doctoral programmes for the European knowledge society', which will collect information about the current situation of doctorates in Europe.

The need for an organized voice of European researchers in the initial stages of their professional career will become increasingly evident in the future, and Eurodoc's growth aims to help meet this challenge. ■

► www.eurodoc.net

Alexandre Urani, Raoul Tan, Renzo Rubele and Daniel Mietchen are members of the Eurodoc board.

MOVERS

Otmar Wiestler, director, German Cancer Research Center (DKFZ), Heidelberg, Germany



If you want to do basic research in modern biomedicine and stay in touch with clinical diagnosis, you have only three fields to choose: human genetics, virology or neuropathology, says Otmar Wiestler. "In other fields the distance to clinical research has either become too great, or the clinical load will be overwhelming, time-wise and mentally."

Wiestler, who on 1 January took over as director of the German Cancer

Research Center (DKFZ), is one such biomedical researcher who has always been keen not to lose sight of the clinical aspects of his work.

Wiestler says that he was lucky as a young scientist to work in groups bristling with talent — one team member in Zürich, Adriano Aguzzi, has since become a leading prion researcher. Wiestler's decision to focus on cancer genes, on the eve of the discovery of the p53 gene as a key tumour suppressor, also proved to be a wise one. All of a sudden, he found himself in the forefront of a newly emerging field.

But he also learned to keep an eye on politics, which proved to be crucial for his future career. He was propelled into the headlines when, three years ago, he and a group leader in Bonn became the first scientists in Germany to apply for public grant money for research involving human embryonic stem cells. The intense public debate that followed

has been a valuable exercise in public relations, he says. But it also had unpleasant side effects: Wiestler's wife and five children were put under police protection when a national newspaper published their home address.

But Wiestler feels no bitterness. He is thrilled by the challenge of transforming the 30-year-old DKFZ into a comprehensive cancer centre to network and coordinate cancer research, diagnosis, prevention and therapy countrywide — similar to the US National Cancer Institute, he says.

Experimental therapies, for example those involving human stem cells, will also play a role at the DKFZ, although not the main role. "It would be very wrong to raise any hopes before their true potential has been thoroughly investigated," he says. "The example of gene therapy has shown us that it can be fatal to test human applications too early." ■

CV

1992–2003: Director of the Institute of Neuropathology, University of Bonn, Germany.

1987–1992: Junior physician, then senior physician and group leader, at the Institute of Pathology, Department of Neuropathology, University Hospital Zurich, Switzerland.

1984–1987: Postdoctoral fellow in the Department of Pathology, University of California, San Diego.

1982–1984: Junior physician in the Department of Neuropathology, University of Freiburg, Germany.